

UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biomediche, Cliniche e Sperimentali

Dipartimento di Scienze Cliniche e di Comunità

Corso di Dottorato di Ricerca in Statistica Biomedica – XXVIII ciclo

Settore scientifico disciplinare MED/01



TESI DI DOTTORATO DI RICERCA

**Metodi statistici per l'analisi e la previsione  
della mortalità per tumore**

Dottoranda: Tiziana Rosso (R10099)

Tutor: Prof. Adriano Decarli

Coordinatore del Dottorato: Prof. Adriano Decarli

A.A. 2014/2015



# Indice

---

|   |           |
|---|-----------|
| <b>Riassunto</b>  | <b>1</b>  |
| <br><b>CAPITOLO 1</b>   |           |
| <b>Introduzione</b>   | <b>5</b>  |
| <br><b>CAPITOLO 2</b>   |           |
| <b>Tassi di mortalità</b>   | <b>7</b>  |
| 2.1 Tassi grezzi di mortalità   | 7         |
| 2.2 Standardizzazione diretta   | 13        |
| 2.3 Distribuzione di probabilità e intervallo di confidenza dei tassi | 15        |
| <br><b>CAPITOLO 3</b>   |           |
| <b>Analisi età-periodo-coorte</b>                                     | <b>18</b> |
| 3.1. Introduzione   | 18        |
| 3.2. Struttura dei dati   | 20        |
| 3.3. Approccio classico   | 22        |
| 3.4. Metodo della verosimiglianza penalizzata                         | 24        |
| 3.5. Metodo dei modelli additivi generalizzati                        | 27        |
| 3.6. Metodo dei minimi quadrati parziali                              | 29        |
| <br><b>CAPITOLO 4</b>   |           |
| <b>Analisi predittiva</b>   | <b>34</b> |
| 4.1. Introduzione   | 34        |
| 4.2. Modello età-periodo con Joinpoint                                | 35        |
| 4.2.1. Introduzione alla Joinpoint                                    | 35        |
| 4.2.2. Joinpoint sul numero di morti                                  | 38        |
| – Modello di regressione lineare                                      | 39        |
| – Modello di regressione di Poisson                                   | 40        |

|   |           |
|---|-----------|
| – Modello di regressione log-lineare              | 41        |
| 4.2.3. Joinpoint sui tassi                        | 42        |
| – Modello di regressione lineare                  | 42        |
| – Modello di regressione di Poisson               | 43        |
| – Modello di regressione log-lineare              | 44        |
| 4.3. Modello età-periodo-coorte bayesiano         | 45        |
| 4.4. Modello età-periodo bayesiano                | 51        |
| 4.5. Test di confronto                            | 53        |
| <br><b>CAPITOLO 5</b>                             |           |
| <b>Applicazione ai dati reali</b>                 | <b>54</b> |
| 5.1. Descrizione dei dati                         | 54        |
| 5.2. Analisi età-periodo-coorte                   | 55        |
| 5.2.1. Risultati                                  | 55        |
| 5.2.2. Confronto dei metodi statistici utilizzati | 58        |
| 5.3. Analisi predittiva                           | 59        |
| 5.3.1. Risultati                                  | 60        |
| 5.3.1.1. Periodi quinquennali                     | 60        |
| 5.3.1.2. Periodi annuali                          | 63        |
| 5.3.2. Confronto dei metodi statistici utilizzati | 66        |
| 5.3.2.1. Periodi quinquennali                     | 66        |
| 5.3.2.2. Periodi annuali                          | 71        |
| <br><b>CAPITOLO 6</b>                             |           |
| <b>Conclusioni</b>                                | <b>76</b> |
| <br><b>Bibliografia</b>                           | <b>79</b> |
| <b>Appendice 1</b>                                | <b>83</b> |
| <b>Appendice 2</b>                                | <b>85</b> |



# Riassunto

---

Nuove tecniche di modellazione per serie storiche hanno reso possibile analizzare i diversi fattori sottostanti ai cambiamenti dei tassi di mortalità e incidenza nel tempo, sia a scopi esplicativi che predittivi. L'analisi età-periodo-coorte contribuisce allo scopo eziologico dell'epidemiologia descrittiva facendo inferenza dal gruppo all'individuo. Questa si riferisce ad una famiglia di tecniche statistiche per comprendere i trend temporali di un outcome, come la mortalità e l'incidenza, in termini di tre variabili temporali: l'età del soggetto, il periodo di calendario e la data di nascita del soggetto (coorte di nascita). Sebbene il modello età-periodo-coorte sia utile, è caratterizzato da un problema strutturale intrinseco di identificabilità: le variabili di età, periodo e coorte hanno una dipendenza lineare esatta, formulabile come "età=periodo-coorte".

Prevedere un evento futuro è un procedimento complesso e insidioso, tuttavia, nella maggior parte delle attività umane, è utile per ottenere informazioni, anche se incerte o imprecise, su probabili andamenti futuri. Prevedere i futuri tassi di incidenza e di mortalità per tumore è necessario per acquisire informazioni utili all'epidemiologia e alla programmazione sanitaria.

Tra i vari metodi per l'analisi età-periodo-coorte che si trovano in letteratura, vengono descritti e confrontati tra loro, attraverso un'applicazione su dati di mortalità provenienti dall'OMS, il metodo della verosimiglianza penalizzata, quello dei modelli additivi generalizzati (GAM) e quello dei minimi quadrati parziali (PLS). Inoltre, vengono presentati e confrontati tra loro, con un'applicazione su dati di mortalità realmente osservati, alcuni metodi di analisi predittiva: alcune tecniche di proiezione della mortalità di breve periodo, che utilizzano un modello età-periodo con joinpoint o un modello età-periodo bayesiano, e una tecnica di lungo periodo, che utilizza un modello età-periodo-coorte bayesiano. In particolare, le previsioni con il modello età-periodo con joinpoint sono effettuate applicando modelli di regressione lineare, Poisson e log-lineare.

Dai confronti eseguiti tra i metodi per l'analisi età-periodo-coorte, si nota che il metodo della verosimiglianza penalizzata e quello dei GAM producono risultati

simili, mentre il metodo dei PLS presenta delle differenze. Possiamo trovare la spiegazione di queste differenze nelle diverse modalità che i tre modelli utilizzano per ovviare al problema della collinearità perfetta dei tre parametri età, periodo e coorte. Da una parte il modello della verosimiglianza penalizzata e quello dei GAM utilizzano tecniche differenti per distribuire l'effetto della deriva temporale tra i periodi e le coorti nel modo più corretto possibile. Dall'altra parte il modello stimato con il metodo dei PLS risolve il problema dell'identificabilità dal punto di vista della risoluzione dell'inversa generalizzata, in modo da minimizzare la matrice di varianze e covarianze tra i possibili parametri stimati. In mancanza di un'analisi di comparazione simulativa formale tra i modelli in oggetto, dobbiamo limitarci a constatare che i due modelli basati sulla distribuzione della deriva sono maggiormente indicati per confronti epidemiologici, dove gli effetti dell'età sono ben definiti (come nel caso della mortalità per tumore) e le problematiche maggiori risiedono in maniera preponderante nello sbrogliare l'effetto deriva dalle coorti e dai periodi. Mentre ci chiediamo se sia possibile che il modello dei PLS si riveli un metodo di predizione dei dati futuri più performante. L'analisi età-periodo-coorte è uno strumento estremamente utile per lo studio dei dati di mortalità, in particolare per lo studio degli effetti di coorte, ma deve essere usato con la dovuta cautela per evitare di essere indotti verso conclusioni sbagliate.

Dal confronto eseguito tra i metodi predittivi considerati, si nota che le previsioni con i diversi modelli sono relativamente simili, specialmente per il modello di Poisson e quello log-lineare. Tuttavia, il modello lineare tende a sottostimare i dati predetti, mentre gli altri tendono a sovrastimare, soprattutto per gli anni di previsione più lontani nel tempo. Complessivamente, il modello bayesiano basato su un'analisi età-periodo sembra risultare meno idoneo a predire la mortalità nel breve-medio periodo rispetto agli altri modelli, che non mostrano grandi differenze di performance tra di loro. Dobbiamo limitarci a constatare che il modello lineare e il modello età-periodo-coorte bayesiano sembrerebbero fornire previsioni della mortalità migliori rispetto a quelle degli altri modelli quando si ha una mortalità bassa, mentre nel caso di un maggiore numero di casi sembrano più performanti i modelli di Poisson e log-lineare. I modelli che predicono meglio dipendono dalla forma della distribuzione sconosciuta dei dati analizzati. Tuttavia, tutti i modelli

considerati risultano essere più appropriati per predire dati di breve periodo. Nessuno dei modelli considerati ha una buona performance per il medio periodo. Bisogna tenere presente che la previsione dei trend futuri è sempre un esercizio complesso e insidioso, e le stime ottenute dovrebbero essere prese con la dovuta cautela e considerate solo come indicazioni generali di potenziale interesse per l'epidemiologia e la programmazione sanitaria.





# CAPITOLO 1

## Introduzione

---

L'epidemiologia descrittiva è vista come un primo approccio alla generazione di ipotesi di ricerca di interesse epidemiologico. Le sue tecniche sono principalmente di natura esplorativa, in quanto il suo scopo principale è quello di generare nuove ipotesi, piuttosto che verificarle.

Storicamente le tecniche di base dell'epidemiologia descrittiva derivano dall'ambito della demografia, dove i tassi di mortalità e morbidità sono i principali strumenti descrittivi e la tecnica della standardizzazione è rimasta per molto tempo l'unico metodo utilizzato per effettuare confronti fra popolazioni diverse, ignorando i problemi di variabilità.

Il miglioramento e la più ampia disponibilità di dati epidemiologici negli anni hanno portato allo sviluppo delle moderne metodologie dell'epidemiologia descrittiva. La raccolta e la diffusione di dati di mortalità ed incidenza, soprattutto di tipo oncologico, è migliorata notevolmente negli anni, specialmente grazie alla diffusione dei registri dei tumori ed agli sforzi fatti per uniformare sia le procedure di raccolta dei dati che la classificazione delle patologie [1]. Inoltre si sono avuti anche miglioramenti riguardanti i dati demografici, i quali sono diventati disponibili per un numero sempre maggiore di popolazioni e sono rilasciati su una base sempre più regolare.

Di conseguenza, si è resa disponibile una grande quantità di serie temporali relative a dati di incidenza e di mortalità, che ha spinto alla creazione di nuove tecniche di modellazione per serie storiche epidemiologiche. Queste hanno reso possibile analizzare i diversi fattori sottostanti ai cambiamenti dei tassi nel tempo, sia a scopi esplicativi che predittivi [2].

Tra i vari metodi e tecniche di modellazione che si trovano in letteratura per studiare i tassi di mortalità, in questa tesi vengono descritti alcuni metodi per l'analisi età-periodo-coorte (*age-period-cohort*, APC), che contribuisce allo scopo eziologico dell'epidemiologia descrittiva facendo inferenza dal gruppo all'individuo. L'analisi

APC si riferisce ad una famiglia di tecniche statistiche per comprendere i trend temporali di un outcome, come l'incidenza o la mortalità del cancro, in termini di tre variabili temporali: l'età del soggetto (*age*, A), il periodo di calendario (*period*, P) e la data di nascita del soggetto (coorte di nascita (*cohort*, C)) [3].

Inoltre, in questa tesi vengono descritti e analizzati alcuni metodi per l'analisi predittiva dei dati di mortalità, che si trovano in letteratura.

Prevedere un evento futuro è un procedimento complesso, soggetto a grandi incertezze e per molti aspetti criticabile. Tuttavia, nella maggior parte delle attività umane, è utile per ottenere informazioni, anche se incerte o imprecise, su probabili andamenti futuri. In demografia, per esempio, è prassi comune produrre proiezioni della struttura della popolazione per i decenni futuri, anche se è noto che i modelli di fertilità possono variare sensibilmente in periodi relativamente brevi, e quindi modificare sostanzialmente la successiva struttura della popolazione.

Prevedere i futuri tassi di incidenza e di mortalità per tumore è necessario per pianificare l'allocazione delle risorse e progettare e valutare le strategie di prevenzione e gestione dei tumori, perché i dati veri sono vecchi di 2-3 o più anni [4].

I metodi per l'analisi APC utilizzati sono stati applicati ai dati di mortalità per tutti i tumori per gli individui di età compresa tra i 20 e i 79 anni, nel periodo compreso tra il 1970 e il 2009, residenti nell'Unione Europea (UE), al fine di mostrarne le caratteristiche e, soprattutto, per acquisire informazioni utili per la ricerca eziologica e la prevenzione di questa neoplasia.

Invece, i metodi per l'analisi predittiva considerati sono stati applicati ai dati di mortalità per i principali siti tumorali (stomaco, intestino, pancreas, polmone, mammella, utero, prostata e leucemia) e per tutti i tumori per gli individui residenti nell'UE, nel periodo compreso tra il 1970 e il 1999, 'prevedendo' i dati per il periodo 2000-2009 e confrontandoli con quelli realmente osservati per tale periodo, per capire quali siano i modelli più performanti.

# CAPITOLO 2

## Tassi di mortalità

---

Tra le varie misure e indicatori sanitari utilizzati in epidemiologia attraverso cui si possono esprimere diversi aspetti dello stato di salute di una popolazione, l'indicatore più usato è il *tasso*. Esso misura il cambiamento istantaneo di una quantità (passaggio da sano a malato o da vivo a deceduto) rispetto alla variazione unitaria di un'altra quantità (tempo). Se l'evento di interesse è rappresentato dalla quantità di morti verificatisi in una popolazione per una data causa in un certo periodo di tempo, l'indicatore più adatto è il *tasso di mortalità*.

### 2.1. Tassi grezzi di mortalità

Il tasso grezzo di mortalità misura la proporzione di individui che muoiono in una popolazione in un certo periodo di tempo e rappresenta il parametro di base per misurare l'andamento della mortalità.

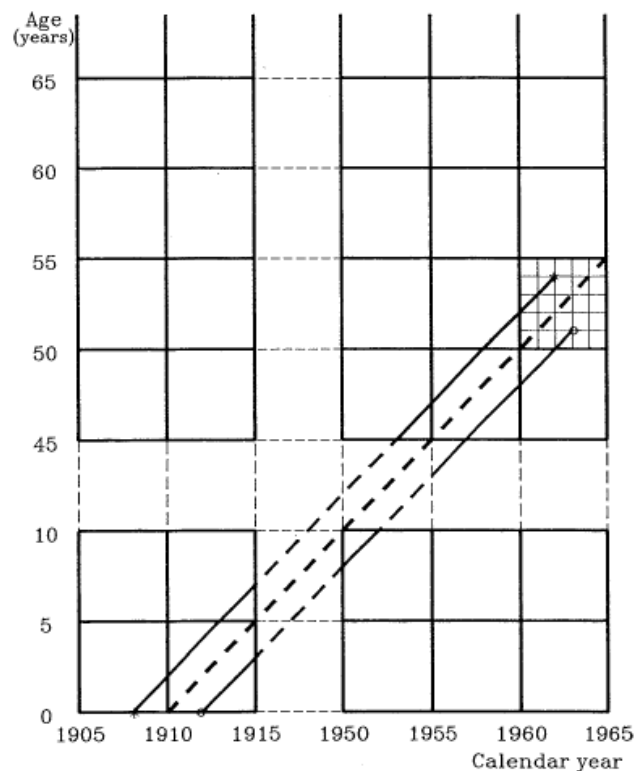
$$\text{Tasso di mortalità} = \frac{\begin{array}{c} \text{n° di morti verificatisi in una} \\ \text{determinata popolazione durante un dato} \\ \text{periodo di tempo} \end{array}}{\begin{array}{c} \text{totale persone – tempo a rischio} \\ \text{durante quel periodo} \end{array}}$$

Il tasso di mortalità esprime la velocità media con la quale un gruppo di individui passa dallo stato di rischio a quello di morte nell'unità di tempo.

Per individuare i fattori che causano un evento (ad es. l'insorgenza di una malattia o il decesso di un individuo) è necessario poterne calcolare il rischio, cioè la probabilità che l'evento si realizzi, la quale dipenderà sia dalle caratteristiche degli individui, come l'età, che dall'ambiente in cui questi sono inseriti.

La funzione di rischio determina come il rischio di sperimentare l'evento cambi nel tempo o con l'età.

Per dare una definizione di rischio è necessario specificare una scala per la misurazione del tempo ed un istante iniziale a partire dal quale verrà misurato il rischio. Il tempo può venire misurato sia rispetto al periodo di calendario che rispetto all'età, corrispondenti ai due assi del diagramma di Lexis [2].



**Figura 1:** Diagramma di Lexis

La struttura del diagramma di Lexis è illustrata in Figura 1. Ogni segmento obliquo rappresenta la frazione della vita di un individuo fra l'inizio dell'osservazione ed il realizzarsi dell'evento in studio. L'estremità sinistra del segmento rappresenta l'inizio dell'osservazione, quella destra il suo termine. Il tempo passato dall'inizio dell'osservazione fino al verificarsi dell'evento può essere misurato proiettando il segmento su uno dei due assi del diagramma.

Fissata una scala temporale, diventa di fondamentale importanza la distribuzione della durata dell'intervallo fra l'inizio dell'osservazione e il verificarsi dell'evento. Se per esempio si è interessati a studiare il rischio di morire in seguito ad una diagnosi di cancro, la conoscenza della distribuzione del tempo intercorrente fra diagnosi e morte permetterebbe di calcolare il rischio di decesso entro un certo anno dalla diagnosi, oppure la probabilità di sopravvivere fino ad un dato anno di età.

Per calcolare effettivamente tali rischi occorre sviluppare un opportuno modello matematico. Il principale assunto è che l'evento in studio sia non ricorrente, ossia una volta verificatosi non possa più ripetersi per lo stesso individuo. Il decesso è chiaramente un evento non ricorrente, mentre l'insorgenza di una malattia non lo è necessariamente.

Per definire il tasso istantaneo o funzione di rischio è necessario prima definire alcune quantità:

- $T$  è una variabile aleatoria rappresentante il tempo d'attesa prima che si verifichi un evento specifico per un individuo appartenente ad una certa popolazione. Inoltre  $T$  deve essere maggiore di zero ( $T > 0$ ), cioè l'individuo non deve ancora aver sperimentato l'evento all'inizio dell'osservazione
- $f(t)$  è la funzione di densità di  $T$ ,  $F(t) = P(T \leq t)$  è la sua distribuzione di probabilità, dove  $F'(t) = f(t)$ , e  $S(t) = P(T > t) = 1 - F(t)$  è la funzione di sopravvivenza, pari alla probabilità che il soggetto sperimenti l'evento oltre un certo istante  $t$ .

Il tasso istantaneo o funzione di rischio è quindi definito come:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t | T \geq t)$$

Esso è dato pertanto dal rapporto tra la probabilità condizionata che si verifichi l'evento al tempo  $t$  del follow-up del soggetto e l'intervallo di tempo  $\Delta t$  associato a tale probabilità [5].

Possiamo notare che il tasso istantaneo  $\lambda(t)$  non è una probabilità, ma una probabilità per unità di tempo, detto anche tasso di probabilità.

Per l'analisi dei dati di mortalità l'evento di interesse è il decesso e di conseguenza la funzione di rischio è anche detta *tasso istantaneo di mortalità*. Tanto più  $\lambda(t)$  è alto

tanto più è probabile che si verifichi un decesso fra  $t$  e l'istante successivo, pertanto  $\lambda(t)$  fornisce una misura della forza di mortalità al tempo  $t$ .

Applicando le regole delle probabilità, è possibile riscrivere  $\lambda(t)$  con la seguente formula:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{F(t + \Delta t) - F(t)}{1 - F(t)}$$

la quale procedendo con i calcoli diventa:

$$\lambda(t)S(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

$$\lambda(t)S(t) = F'(t)$$

Così la probabilità che l'evento si verifichi prima di un certo tempo  $t$ ,  $\pi(t)$ , può essere scritta come segue:

$$\pi(t) = \int_0^t F'(u) du$$

o anche:

$$\pi(t) = \int_0^t \lambda(u)S(u) du$$

Si arriva quindi a definire la probabilità  $\pi$  di morire tra l'età  $t_0$  e l'età  $t_1$  nel seguente modo:

$$\pi = \int_{t_0}^{t_1} \lambda(u)S(u) du$$

con  $\lambda(u)$  tasso specifico per età e  $S(u)$  probabilità di sopravvivere [2].

L'interesse primario risulta rivolto al calcolo della probabilità condizionata  $\pi_c$  di morire tra l'età  $t_0$  e  $t_1$  dato un soggetto ancora a rischio all'età  $t_0$ . Questa probabilità

non è influenzata dalla sopravvivenza generale fino all'età esaminata  $t_0$  e molto poco influenzata dalla sopravvivenza tra  $t_0$  e  $t_1$  se questo intervallo è piccolo.

Quindi è possibile calcolare  $\pi_c$  tramite la formula:

$$\pi_c = \int_{t_0}^{t_1} \lambda(u) \frac{S(u)}{S(t_0)} du$$

Sotto l'assunto che l'intervallo  $[t_0, t_1]$  sia sufficientemente piccolo così che  $\lambda(u)$  e  $S(u)$  possono essere considerati costanti ( $\lambda(t_0)$  e  $S(t_0)$ ), allora l'equazione può essere riscritta nella seguente forma:

$$\pi_c \approx \lambda(t_0)(t_1 - t_0)$$

Definito  $d$  il numero di morti osservate tra  $t_0$  e  $t_1$ , e  $n_{t_0}$  il numero di soggetti a rischio a  $t_0$ , allora la stima di  $\pi_c$  è data da:

$$\widehat{\pi}_c = \frac{d}{n_{t_0}}$$

e la stima di  $\lambda$  è:

$$\hat{\lambda}(t_0) \approx \frac{d}{n_{t_0}(t_1 - t_0)}$$

Di conseguenza la stima del tasso istantaneo a  $t_0$  è ottenuta dividendo il numero di morti osservate per il numero  $m$  di anni-persona tra  $t_0$  e  $t_1$ , che può essere riscritta nella forma più familiare:

$$\hat{\lambda}(t_0) \approx \frac{d}{m}$$

E' da sottolineare che questa approssimazione non risulta adatta quando  $\lambda(u)$  varia fortemente tra  $t_0$  e  $t_1$ , o se il rapporto  $S(u)/S(t_0)$  è molto diverso dall'unità [2].

Esplicitando con parole i termini considerati si arriva alla seguente definizione del tasso istantaneo [5]:



$$\lambda = \frac{p}{t} = \frac{d/n}{t} = \frac{d}{n * t}$$

$$\lambda = \frac{\text{probabilità}}{t} = \frac{\text{morti}/\text{persone a rischio}}{\text{tempo di osservazione}} =$$

$$= \frac{\text{morti}}{\text{persone a rischio} * \text{tempo di osservazione}}$$

Concludendo, il tasso risulta essere una grandezza concettualmente istantanea, in quanto il periodo di osservazione costituisce solo l'elemento per poter effettuare il calcolo in pratica (tasso medio). Inoltre, proprio perché il tempo di osservazione compare direttamente nella definizione della grandezza, il tasso tiene conto non solo delle morti avvenute ma anche del tempo in cui sono avvenute [5].

Nel caso dei tassi di mortalità annui, il numeratore è rappresentato dal numero di morti per la condizione oggetto di indagine nel corso dell'anno solare, mentre il denominatore, nel caso delle statistiche descrittive, è costituito da stime della popolazione derivanti dai censimenti. In altri termini gli anni-persona considerati per ogni classe di età sono espressi come il numero di individui presenti a metà anno. In formula:

$$\hat{\lambda}_x = \frac{d_x}{[n_x(t_0) + n_x(t_1)]/2}$$

dove:

$\hat{\lambda}_x$  è la stima del tasso di mortalità annuale per la classe di età studiata,

$d_x$  è il numero di decessi osservati nella popolazione in studio,

$n_x(t_0)$  e  $n_x(t_1)$  sono rispettivamente il numero di soggetti a rischio all'inizio e alla fine del periodo in studio.

In questa formula, l'espressione a denominatore corrisponde al valore  $m_x$  presente nelle altre formule di questa tesi.

## 2.2. Standardizzazione diretta

Il tasso presentato nel precedente paragrafo è un tasso grezzo, mentre per poter confrontare i livelli di mortalità fra diverse regioni geografiche o periodi di calendario è necessario tenere in considerazione quei fattori che possono agire da confondenti, ricorrendo al calcolo dei tassi standardizzati. La distribuzione per età degli individui può cambiare fra le popolazioni confrontate e poiché il verificarsi di molte condizioni di salute è connesso all'età, la standardizzazione più comune per i dati riguardanti la salute pubblica è la standardizzazione per età.

Esistono due metodi di standardizzazione dei tassi: diretta e indiretta, ed in entrambi i casi si può standardizzare per diverse variabili. Poiché per confrontare l'incidenza o la mortalità tra due o più popolazioni si ottengono stime meno distorte con la standardizzazione diretta che con quella indiretta, in questa tesi faremo riferimento alla standardizzazione diretta per età [6].

La standardizzazione diretta per età consiste nel determinare quale sarebbe il tasso annuo in una popolazione di riferimento (popolazione standard), con una data struttura per età, se questa fosse soggetta alla stessa mortalità della popolazione in studio. La popolazione di riferimento non è necessariamente una popolazione reale, può anche essere una popolazione teorica.

In altre parole, si stimano il numero atteso di morti in ogni classe di età nella popolazione standard, applicando i corrispondenti tassi specifici della popolazione in esame agli anni-persona della popolazione standard. Successivamente il numero totale di decessi attesi calcolati viene diviso per il totale degli anni-persona della popolazione standard.

Pertanto il *tasso standardizzato* è dato dalla seguente formula:

$$\lambda_{SD} = \frac{1}{L} \sum_{x=1}^g L_x \lambda_x$$

dove:  $g$  = numero totale di classi di età

$L_x$  = numerosità della popolazione standard nell' $x$ -esima classe di età

$\lambda_x$  = tasso nell' $x$ -esima classe di età per la popolazione in esame

$L$  = numerosità totale della popolazione standard

che può anche essere scritta come:

$$\lambda_{SD} = \sum_{x=1}^g w_x \lambda_x$$

dove:  $w_x = \frac{L_x}{L}$  è la proporzione di individui nell' $x$ -esima classe di età della popolazione standard, per la quale vale la seguente regola:

$$\sum_{x=1}^g w_x = \frac{1}{L} \sum_{x=1}^g L_x = 1$$

Quindi  $\lambda_{SD}$  è una media ponderata dei tassi età-specifici, con i pesi che equivalgono alla proporzione di individui per classe di età nella popolazione standard [7].

Per quanto concerne la scelta e la definizione della popolazione di riferimento (popolazione standard) esistono vari metodi. Il metodo più semplice è quello di definire a priori una popolazione fittizia da usare come standard: la popolazione mondiale standard, a cui si farà riferimento in questa analisi, è una di queste popolazioni teoriche. Essa è costruita dall'OMS (*Organizzazione Mondiale della Sanità*) per facilitare i confronti internazionali e la sua struttura per età è riportata in Tabella 1 [7].

| Classe di età | Popolazione standard |
|---------------|----------------------|
| 0-4           | 12                   |
| 5-9           | 10                   |
| 10-14         | 9                    |
| 15-19         | 9                    |
| 20-24         | 8                    |
| 25-29         | 8                    |
| 30-34         | 6                    |

|        |     |
|--------|-----|
| 35-39  | 6   |
| 40-44  | 6   |
| 45-49  | 6   |
| 50-54  | 5   |
| 55-59  | 4   |
| 60-64  | 4   |
| 65-69  | 3   |
| 70-74  | 2   |
| 75-79  | 1   |
| 80-84  | 0.5 |
| 85 +   | 0.5 |
| Totale | 100 |

**Tabella 1:** Pesi delle classi di età della popolazione mondiale standard.

### 2.3. Distribuzione di probabilità e intervallo di confidenza dei tassi

La distribuzione di probabilità esatta per il tasso risulta molto complicata a causa della presenza dei dati censurati [5]. Per ogni ipotetica unità  $i$  osservata per il tempo  $T_i$  si è identificato che gli eventi sono distribuiti adeguatamente secondo la legge di probabilità di Poisson; poiché gli eventi hanno una probabilità costante di realizzarsi nel tempo, il succedersi di un evento non influenza il verificarsi dell'evento successivo e la probabilità che due eventi si verifichino nello stesso istante è nulla. Se si estende questo ragionamento a tutto lo studio si può assumere di avere tanti piccoli processi poissoniani. Ipotizzando che tutte le unità sperimentali siano indipendenti (come nel caso della mortalità), lo studio può essere quindi pensato come un insieme di processi poissoniani indipendenti: il risultato sarà una distribuzione di Poisson relativa al totale degli eventi [5].

In formula:

$$pr(d = x|D) = \frac{e^{-D} \cdot D^x}{x!} = \frac{e^{-\lambda m} (\lambda m)^x}{x!}$$

dove:

$D$  è il totale dei decessi,

$\lambda$  è il tasso di mortalità,

$m$  è il totale delle persone-tempo sotto osservazione.

Quindi  $E(D) = \lambda m$  e  $Var(D) = \lambda m$  [5, 8].

Se  $\lambda$  è sufficientemente piccolo e  $D$  è molto più piccolo di  $m$  la distribuzione di Poisson costituisce una buona approssimazione della distribuzione probabilistica esatta di un tasso. Inoltre la variabilità del tasso è considerata associata solo al suo numeratore (gli eventi osservati), mentre il denominatore (la popolazione-tempo) è considerato fisso e quindi non affetto da variabilità casuale. Pertanto l'accuratezza di un tasso dipende solo dalla variabilità del numero dei casi osservati ( $D$ ) [2]. Quando gli eventi non sono troppo pochi la distribuzione di probabilità del tasso è approssimativamente gaussiana con media  $\mu = D/m$  e con varianza  $\sigma^2 = D/m^2$  [5]. Pertanto con questa approssimazione, lo stimatore del tasso ( $\hat{\lambda} = D/m$ ) ha varianza pari a:

$$Var(\hat{\lambda}) = Var\left(\frac{D}{m}\right) = \frac{Var(D)}{m^2} = \frac{\lambda}{m}$$

Da qui osserviamo che la varianza del tasso specifico  $\hat{\lambda}_x$  è:

$$Var(\hat{\lambda}_x) = \frac{Var(D_x)}{m_x^2} = \frac{\lambda_x}{m_x}$$

Per il tasso standardizzato diventa:

$$Var(\hat{\lambda}_{SD}) = \sum_{x=1}^g w_x^2 Var(\hat{\lambda}_x) = \sum_{x=1}^g w_x^2 \left(\frac{\lambda_x}{m_x}\right)$$

Poiché  $\lambda$  è non noto, questa varianza deve essere stimata con [2]:

$$\widehat{Var}(\hat{\lambda}_{SD}) = \sum_{x=1}^g \left(\frac{w_x^2}{m_x^2}\right) d_x$$

Se il tasso standardizzato teorico è dato da:

$$\lambda_{SD} = \sum_x w_x \lambda_x$$

e se  $s$  è la stima del suo errore standard, allora  $(\hat{\lambda}_{SD} - \lambda_{SD})/s$  è approssimativamente una variabile normale standard; è possibile allora calcolare l'intervallo di confidenza (IC) approssimato con la consueta formula:

$$IC_{1-\alpha}(\lambda_{SD}) = \left[ \hat{\lambda}_{SD} - z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\lambda}_{SD})}, \hat{\lambda}_{SD} + z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\lambda}_{SD})} \right]$$

dove  $1 - \alpha$  è il livello di confidenza prescelto e  $z_{1-\frac{\alpha}{2}}$  è il quantile di livello  $1 - \frac{\alpha}{2}$  di una normale standard  $N(0,1)$ .

# CAPITOLO 3

## Analisi età-periodo-coorte

---

### 3.1. Introduzione

I tassi standardizzati sono un utile strumento descrittivo quando si confrontano diverse popolazioni, ma le informazioni prodotte non hanno carattere analitico; d'altra parte, metodi analitici inferenziali, applicati ai trend temporali dei tassi di mortalità come l'analisi joinpoint, permettono di analizzare i trend e il loro cambiamento nel tempo, ma tralasciano formalmente le informazioni sulla struttura di età e di coorte dei dati analizzati [9].

L'importanza di distinguere tra gli effetti di età, periodo e coorte di nascita nell'analisi dell'incidenza e della mortalità di una malattia è stata riconosciuta già negli anni '30, quando il problema venne affrontato attraverso un approccio descrittivo che analizzava separatamente i tre effetti [10]. Sebbene tali tecniche fornivano maggiori informazioni rispetto all'analisi trasversale standard dei tassi di mortalità, risultavano comunque non del tutto soddisfacenti. Per questa ragione, negli anni sono stati proposti e implementati modelli statistici per distinguere gli effetti della coorte di nascita, del periodo di morte e dell'età [11, 12].

L'*analisi età-periodo-coorte* (APC) studia simultaneamente gli effetti dell'età, del periodo e della coorte di nascita [13].

L'età del soggetto al momento dell'evento si riferisce al tempo trascorso dalla nascita, mentre il periodo si riferisce alla data di calendario in cui si è verificato l'evento di interesse. Infine, la coorte di nascita identifica il tempo di calendario di quando un individuo è nato.

Lo scopo dell'analisi età-periodo-coorte è di determinare i contributi separati dell'età, del periodo e della coorte di nascita sull'evento in studio [3].

Gli effetti dell'età corrispondono alla variabilità spiegata dalle differenze fisiologiche che caratterizzano i diversi gruppi di età. Gli effetti di periodo sono provocati dal

cambiamento di variabilità nel tempo che influenza tutte le età allo stesso modo, influiscono su tutti gli individui sotto osservazione a una certa data, indipendentemente dalla loro età; solitamente descrivono le conseguenze dell'introduzione di nuove terapie o interventi di screening e sono anche influenzati da esposizioni che producono conseguenze sull'evento studiato nel breve periodo. Gli effetti di coorte, invece, spiegano i cambiamenti che interessano i gruppi di soggetti appartenenti alla stessa coorte di nascita; di solito sono determinati da esposizioni gestazionali, o da quelle esposizioni che influenzano la frequenza dell'evento nel lungo periodo [3, 14].

Sebbene il modello APC sia utile, è caratterizzato da un problema strutturale intrinseco: le variabili di età, periodo e coorte hanno una dipendenza lineare esatta, formulabile come "età=periodo-coorte ( $A=P-C$ )". Questo causa un *problema di identificabilità* che rende questo modello complicato da trattare [15]. Negli ultimi decenni, molti sforzi sono stati fatti nel tentativo di superare questo problema sia ricorrendo ad approcci di tipo parametrico che non parametrico [16]. Alcuni esempi sono l'utilizzo di funzioni stimabili, l'aggiunta di vincoli supplementari al modello o l'uso di un approccio a cascata in due fasi, che esegue adattamenti sequenziali di un modello a due fattori e poi regredisce la variabile rimanente sui residui [13, 16-20]. Più recentemente, si è provato a risolvere il problema anche attraverso l'introduzione di tecniche coinvolgenti metodi GAM (*Generalised Additive Models*), *smoothers*, Bayesiani e PLS (*Partial Least Squares*) [16, 17, 21-24].

Ciascuna di queste soluzioni presenta i propri vantaggi e svantaggi: le funzioni stimabili sono epidemiologicamente difficili da interpretare; l'aggiunta di vincoli richiede una conoscenza biologica e statistica a priori e non è sempre giustificabile; le stime della regressione a cascata non sempre sono concordi tra loro e richiedono anche la giustificabilità biologica.

Nei seguenti paragrafi procederemo nella presentazione dell'analisi APC, illustrando alcuni di questi metodi al fine di sottolinearne i punti di forza e gli eventuali limiti.



### 3.2. Struttura dei dati

I dati di incidenza e di mortalità di una popolazione per una determinata causa sono spesso mostrati attraverso una tabella di contingenza a due vie, in cui le righe rappresentano le classi di età con le quali è stata raggruppata la popolazione ( $i = 1, \dots, I$ ) e le colonne i periodi di tempo, usualmente, ma non necessariamente, di ampiezza pari a quella delle classi di età, in cui la stessa popolazione è stata osservata ( $j = 1, \dots, J$ ). Ogni cella può contenere una o più delle seguenti informazioni:

- il numero di eventi avvenuti ( $O_{ij}$ ) nei soggetti appartenenti alla classe di età  $i$  e nel corso del periodo  $j$ , che corrisponde al numero di decessi nel caso della mortalità o ai nuovi casi di malattia nel caso dell'incidenza;
- il numero di anni-persona a rischio ( $N_{ij}$ ), dato dalla somma del numero di anni in cui ogni soggetto appartenente alla classe  $i$  è stato osservato nel corso del periodo  $j$ ;
- il tasso specifico per età ( $r_{ij}=O_{ij}/N_{ij}$ ), che sintetizza le informazioni sopra specificate.

Le Tabelle 2, 3 e 4 rappresentano un tipico esempio di tabelle di mortalità per tumore del polmone per i maschi in Europa, in cui le classi di età (per riga) e i periodi di tempo (per colonna) sono raggruppati in quinquenni e le celle contengono rispettivamente i corrispondenti numero di morti (Tabella 2), anni-persona a rischio (Tabella 3) e tassi specifici per età per 100.000 abitanti (Tabella 4). Le coorti di nascita, invece, possono essere direttamente ricavate dalla tabella; la coorte corrispondente all' $i$ -esima classe di età e al  $j$ -esimo periodo di morte è rappresentata da  $k=I-i+j$ . Si può notare che scorrendo verticalmente la tabella verso il basso si osserva la variazione sulle classi di età ( $I=12$ ), spostandosi verso destra lungo le colonne si analizza invece la variazione temporale sui periodi quinquennali ( $J=8$ ); spostandosi in diagonale verso il basso si hanno le variazioni attraverso le coorti ( $K=I+J-1=19$ ). Per semplificare la comprensione, nella tabella si è evidenziata la coorte con anno di nascita centrale nel 1950 ( $C=P-A$ ).

In questo esempio (Tabella 4) si può notare facilmente come i tassi di mortalità aumentino al crescere dell'età e verso i primi periodi di morte; la dipendenza

intrinseca degli effetti di coorte dalle variabili di età e periodo appare chiaramente anche in questo esempio.

|              | Periodo quinquennale di morte |         |         |         |         |         |         |         |
|--------------|-------------------------------|---------|---------|---------|---------|---------|---------|---------|
|              | 1970-74                       | 1975-79 | 1980-84 | 1985-89 | 1990-94 | 1995-99 | 2000-04 | 2005-09 |
| <b>20-24</b> | 304                           | 288     | 223     | 228     | 183     | 142     | 113     | 24      |
| <b>25-29</b> | 535                           | 556     | 527     | 439     | 388     | 339     | 260     | 52      |
| <b>30-34</b> | 1241                          | 1398    | 1494    | 1555    | 1326    | 994     | 852     | 133     |
| <b>35-39</b> | 3806                          | 4258    | 4306    | 5058    | 5034    | 4279    | 3071    | 494     |
| <b>40-44</b> | 11191                         | 11657   | 12261   | 12908   | 14800   | 13830   | 11184   | 1702    |
| <b>45-49</b> | 24677                         | 28913   | 29060   | 30081   | 30122   | 32038   | 29618   | 4879    |
| <b>50-54</b> | 40820                         | 54218   | 60678   | 59337   | 57072   | 54529   | 57241   | 10997   |
| <b>55-59</b> | 65830                         | 73977   | 96959   | 104678  | 97620   | 88702   | 84009   | 18327   |
| <b>60-64</b> | 116021                        | 104881  | 112788  | 142297  | 146344  | 130044  | 119035  | 23008   |
| <b>65-69</b> | 142647                        | 151731  | 131996  | 141630  | 171156  | 169658  | 151869  | 28754   |
| <b>70-74</b> | 123425                        | 148513  | 157299  | 137063  | 145345  | 170422  | 170786  | 31702   |
| <b>75-79</b> | 70271                         | 98409   | 119532  | 128564  | 112740  | 121373  | 145028  | 31035   |

**Tabella 2:** Numero di morti per tumore del polmone per i maschi in Europa suddivisi per classe quinquennale di età e periodo quinquennale di morte.

|              | Periodo quinquennale di morte |          |          |          |          |          |          |          |
|--------------|-------------------------------|----------|----------|----------|----------|----------|----------|----------|
|              | 1970-74                       | 1975-79  | 1980-84  | 1985-89  | 1990-94  | 1995-99  | 2000-04  | 2005-09  |
| <b>20-24</b> | 87342250                      | 88706650 | 91651743 | 94063989 | 91543654 | 85242294 | 82020679 | 16182715 |
| <b>25-29</b> | 76324200                      | 86047600 | 87545111 | 90962498 | 94243523 | 91964611 | 86004318 | 17094738 |
| <b>30-34</b> | 73959750                      | 75818750 | 85083847 | 86785846 | 91001756 | 94397670 | 91981262 | 17677109 |
| <b>35-39</b> | 73126150                      | 73116050 | 74911154 | 84055352 | 86468093 | 90509257 | 93694743 | 18805350 |
| <b>40-44</b> | 71414150                      | 70991550 | 71919913 | 73848233 | 83481703 | 85584570 | 89514263 | 18870629 |
| <b>45-49</b> | 67069050                      | 69590750 | 69466621 | 70404883 | 72498104 | 81800686 | 84093420 | 17782926 |
| <b>50-54</b> | 55426250                      | 64190850 | 67091849 | 67093019 | 68497830 | 70428310 | 79446083 | 16458640 |
| <b>55-59</b> | 48129950                      | 51130850 | 60554288 | 63507670 | 63828606 | 65235952 | 67445538 | 15429372 |
| <b>60-64</b> | 52059350                      | 45016850 | 46809705 | 55615055 | 58739329 | 59324430 | 61203802 | 12624770 |
| <b>65-69</b> | 43343350                      | 45355850 | 39104830 | 41087421 | 49171166 | 52149438 | 53716505 | 11262800 |
| <b>70-74</b> | 30836650                      | 34557950 | 36467333 | 31625081 | 34065859 | 41158787 | 44490810 | 9347067  |
| <b>75-79</b> | 17649400                      | 21120300 | 24155603 | 25981923 | 23462009 | 25793450 | 31835101 | 7167792  |

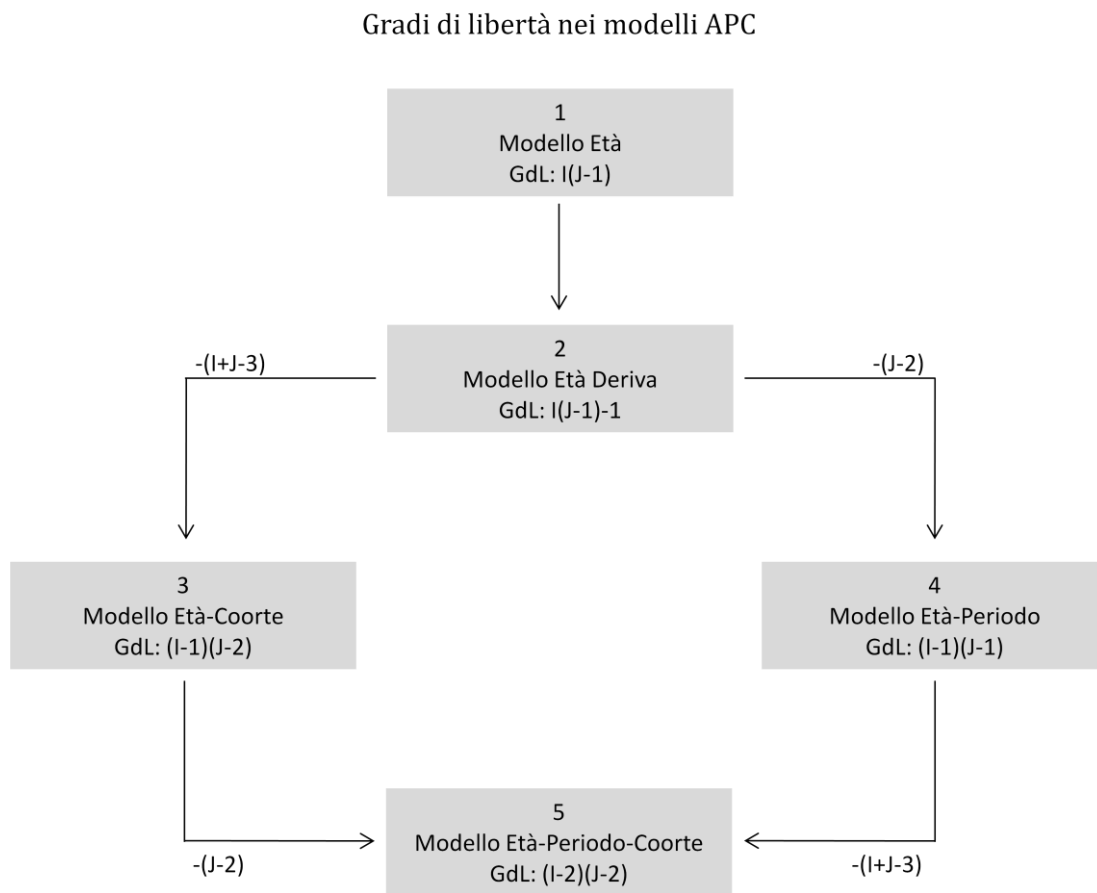
**Tabella 3:** Anni-persona a rischio per i maschi in Europa suddivisi per classe quinquennale di età e periodo quinquennale di morte.

|               |       | Periodo quinquennale di morte |         |         |         |         |         |         |         |
|---------------|-------|-------------------------------|---------|---------|---------|---------|---------|---------|---------|
|               |       | 1970-74                       | 1975-79 | 1980-84 | 1985-89 | 1990-94 | 1995-99 | 2000-04 | 2005-09 |
| Classe di età | 20-24 | 0,35                          | 0,32    | 0,24    | 0,24    | 0,20    | 0,17    | 0,14    | 0,15    |
|               | 25-29 | 0,70                          | 0,65    | 0,60    | 0,48    | 0,41    | 0,37    | 0,30    | 0,30    |
|               | 30-34 | 1,68                          | 1,84    | 1,76    | 1,79    | 1,46    | 1,05    | 0,93    | 0,75    |
|               | 35-39 | 5,20                          | 5,82    | 5,75    | 6,02    | 5,82    | 4,73    | 3,28    | 2,63    |
|               | 40-44 | 15,67                         | 16,42   | 17,05   | 17,48   | 17,73   | 16,16   | 12,49   | 9,02    |
|               | 45-49 | 36,79                         | 41,55   | 41,83   | 42,73   | 41,55   | 39,17   | 35,22   | 27,44   |
|               | 50-54 | 73,65                         | 84,46   | 90,44   | 88,44   | 83,32   | 77,42   | 72,05   | 66,82   |
|               | 55-59 | 136,77                        | 144,68  | 160,12  | 164,83  | 152,94  | 135,97  | 124,56  | 118,78  |
|               | 60-64 | 222,86                        | 232,98  | 240,95  | 255,86  | 249,14  | 219,21  | 194,49  | 182,24  |
|               | 65-69 | 329,11                        | 334,53  | 337,54  | 344,70  | 348,08  | 325,33  | 282,72  | 255,30  |
|               | 70-74 | 400,25                        | 429,75  | 431,34  | 433,40  | 426,66  | 414,06  | 383,87  | 339,17  |
|               | 75-79 | 398,15                        | 465,94  | 494,84  | 494,82  | 480,52  | 470,56  | 455,56  | 432,98  |

**Tabella 4:** Tassi di mortalità per 100.000 per tumore del polmone per i maschi in Europa suddivisi per classe quinquennale di età e periodo quinquennale di morte.

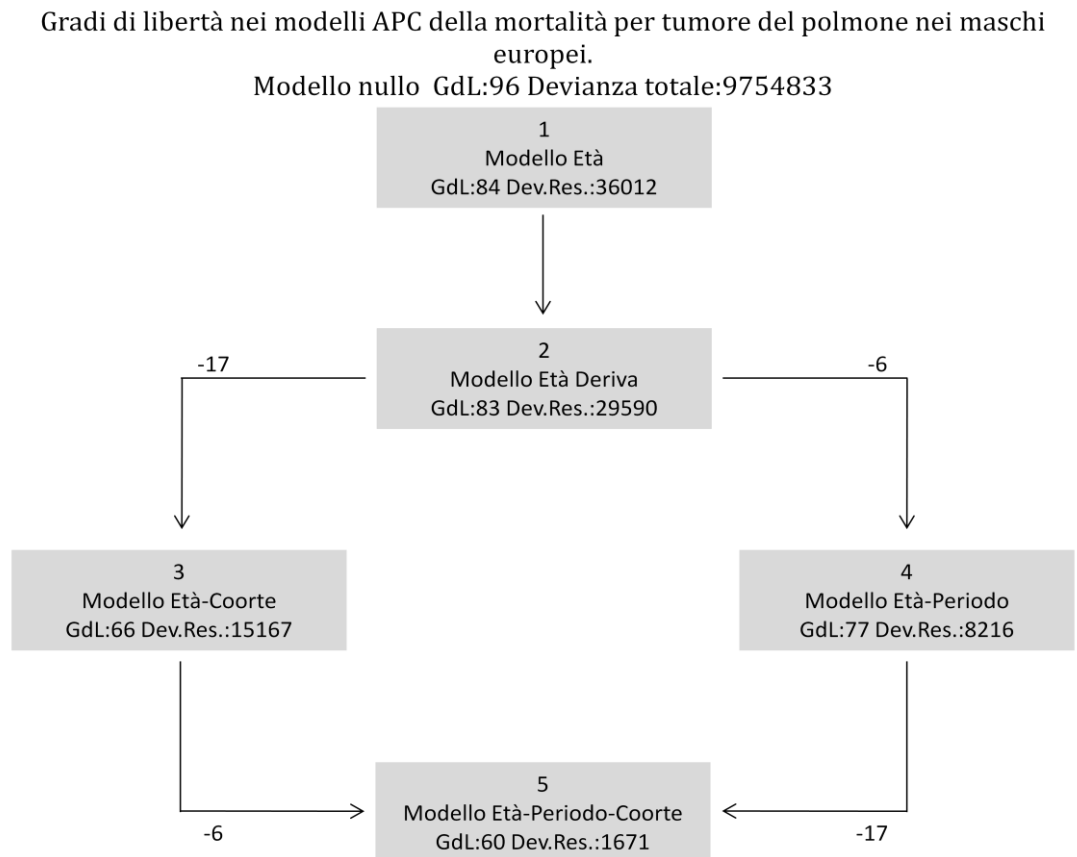
### 3.3. Approccio classico

Per illustrare i metodi e i problemi che derivano dalla applicazione dei modelli APC è opportuno ricorrere alla trattazione classica che Clayton e Schifflers hanno esposto nel loro doppio articolo intitolato “Models for Temporal Variation in Cancer Rates” [18, 25]. Nel lavoro, gli autori affrontano il problema eseguendo una serie di analisi della varianza (*Analysis of Variance*, ANOVA) di modelli progressivamente più complessi. Partirono da un semplice modello età, passando a un modello di deriva lineare che aggiunge un termine di deriva di coorte o di periodo al modello di età; applicarono successivamente i modelli età-periodo (AP) ed età-coorte (AC) e solo in ultimo, se giustificato dall’analisi della varianza, applicarono una delle possibili soluzioni APC. Il diagramma di flusso presentato nel loro articolo (Figura 2), rappresenta sinteticamente la sequenza del metodo.



**Figura 2:** Schema della procedura ANOVA per l'analisi età-periodo-coorte.

Questo schema sottolinea come i modelli AP e AC non siano direttamente confrontabili, poiché il modello AC usa più gradi di libertà del modello AP. Di conseguenza l'ANOVA risultante è complicata da interpretare. Applicando questo metodo ANOVA ai modelli sopra citati in riferimento ai tassi di mortalità per tumore del polmone dei maschi europei del precedente esempio otteniamo:



**Figura 3:** Schema ANOVA applicato alla mortalità per tumore del polmone dei maschi europei.

Esaminando la Figura 3, si può vedere che il modello età deriva spiega una quantità significativa di variabilità utilizzando un solo grado di libertà, quindi dà la rappresentazione più economica dei dati.

### 3.4. Metodo della verosimiglianza penalizzata

I modelli costruiti per analizzare i tassi di mortalità sono modelli additivi dei parametri delle variabili di interesse. Essi hanno come variabile dipendente la variabile  $Y = \log(E[r])$ , pari alla trasformazione logaritmica del tasso atteso  $r$ . I tassi sono approssimativamente distribuiti secondo la legge di probabilità di Poisson, così come gli eventi osservati. Dal momento che il logaritmo è una trasformazione monotona crescente, allora il modello che verrà utilizzato in seguito può essere denominato modello log-lineare di Poisson.

Un modello APC completo è scritto come:

$$Y = \log(E[r_{ijk}]) = \alpha_i + \beta_j + \gamma_k \quad (1)$$

Per stimare questo modello è necessario prima risolvere o aggirare il problema della identificabilità. Una soluzione per risolvere tale difficoltà consiste nel usare una funzione di verosimiglianza penalizzante la media dei 3 possibili modelli a due fattori [12].

Un modello APC può essere considerato come un problema di stima dei parametri in un modello log-lineare di Poisson:

$$\log(E[o_{ijk}]) = \log(N_{ijk}) + \log a_i + \log p_j + \log c_k$$

dove  $\log a_i$ ,  $\log p_j$  e  $\log c_k$  corrispondono rispettivamente a  $\alpha_i$ ,  $\beta_j$  e  $\gamma_k$  della formula (1), ovvero  $a_i$  ( $i=1, \dots, I$ ),  $p_j$  ( $j=1, \dots, J$ ) e  $c_k$  ( $k=1, \dots, K$ ) sono i parametri moltiplicativi rispettivamente per l'età, il periodo e la coorte. Tali parametri possono essere stimati con il metodo dei minimi quadrati pesati, cioè minimizzando la seguente funzione:

$$f(a, p, c) = \sum o_{ijk} (\log o_{ijk} - \log N_{ijk} - \log a_i - \log p_j - \log c_k)^2 \quad (2)$$

Dato che c'è una relazione lineare tra A, P e C l'insieme delle soluzioni  $X(a, p, c)$  è infinito. Per risolvere questo problema, l'insieme delle soluzioni può essere parametrizzato in  $\lambda$  come  $X(\lambda) = (a, p, c, \lambda)$  dato da:

$$\log \hat{a}'_i = \log \hat{a}_i + \lambda(I - i)$$

$$\log \hat{p}'_j = \log \hat{p}_j + \lambda j$$

$$\log \hat{c}'_k = \log \hat{c}_k - \lambda k$$

Questa parametrizzazione rende possibile calcolare una statistica della bontà di adattamento ( $G^2$ ) che è indipendente da  $\lambda$ .

Una soluzione al problema di identificabilità è stata presentata nel 1982 da Osmond e Gardner [11] calcolando tre modelli log-lineari a due fattori, che minimizzano rispettivamente le seguenti tre funzioni di tipo (2):

$$f(a, p, c_0) ; f(a, p_0, c) ; f(a_0, p, c)$$

dove  $c_0$  e  $p_0$  sono vettori unitari rispettivamente di lunghezza K e J e  $a_0$  è un vettore pari a:

$$a_{oi} = \exp \left[ \frac{\sum_j O_{ij} (\log O_{ij} - \log N_{ij})}{\sum_j O_{ij}} \right]$$

Da questi calcoli si ottengono come soluzioni i seguenti vettori delle stime dei parametri:

$$X_c = (\hat{a}_c, \hat{p}_c, c_0); X_p = (\hat{a}_p, p_0, \hat{c}_p); X_a = (a_0, \hat{p}_a, \hat{c}_a)$$

Ciascuno dei tre modelli è ovviamente caratterizzato da una propria misura della bontà di adattamento ( $G_c^2$ ,  $G_p^2$  e  $G_a^2$ ).

Considerando il logaritmo naturale di queste soluzioni nello spazio reale  $R^{i+j+k}$ , le tre distanze euclidee tra ciascun modello a due fattori e il modello saturo parametrizzato in  $\lambda$ ,  $X(\lambda)$ , sono definite come:

$$d_c(\lambda) = \|X_c - X(\lambda)\|$$

$$d_p(\lambda) = \|X_p - X(\lambda)\|$$

$$d_a(\lambda) = \|X_a - X(\lambda)\|$$

La somma di queste distanze, pesata per l'inverso delle misure della bontà di adattamento divise per i rispettivi gradi di libertà, e cioè:

$$g(X) = \frac{d_a(\lambda)}{(G_a^2/(I-2)(J-1))} + \frac{d_p(\lambda)}{(G_p^2/(I-1)(J-2))} + \frac{d_c(\lambda)}{(G_c^2/(I-1)(J-1))}$$

può essere minimizzata rispetto a  $\lambda$ . In altre parole, questo metodo tra tutte le possibili soluzioni  $X(\lambda)$  sceglie quella corrispondente ad un valore  $X'(\lambda)$  tale per cui  $g(X)$  risulti minima. Quindi  $X'(\lambda)$  rappresenta il vettore finale delle stime degli effetti del modello età-periodo-coorte. Questa soluzione minimizza le distanze dei modelli saturi dai tre modelli a due fattori e costruisce una media geometrica ponderata dei tre modelli a due fattori [12]. In questo modo il problema dell'identificabilità è risolto e la deriva è distribuita secondo le statistiche della bontà di adattamento.

Molti metodi dell'analisi APC non consentono di calcolare gli intervalli di confidenza in modo convenzionale; tra questi troviamo il metodo della verosimiglianza penalizzata. Pertanto, per ottenere informazioni sulla variabilità dei parametri stimati, si ricorre ad un metodo parametrico di simulazione bootstrap [26]. Solitamente i dati di mortalità pubblicati sono stratificati per classi di età e periodi. Il numero di morti presenti in ogni cella, specifici per ogni classe di età e periodo, sono

distribuiti secondo la legge di probabilità di Poisson. Tenendo presente questa considerazione, per ogni cella, si estraggono casualmente 1000 valori da una distribuzione di Poisson caratterizzata dal valore misurato in quella cella. Gli insiemi di dati simulati sono poi gestiti attraverso il metodo APC della verosimiglianza penalizzata, ottenendo così 1000 insiemi di stime per i parametri dell'età, del periodo e della coorte.

Per ogni parametro il 2,5-esimo e il 97,5-esimo percentile sono presi dall'insieme di mille stime come stima per l'intervallo di confidenza al 95% per quel parametro.

Per l'applicazione mostrata in questa tesi sono state effettuate 1000 simulazioni di questo modello.

### 3.5. Metodo dei modelli additivi generalizzati

Un altro approccio per la modellazione età-periodo-coorte è quello di considerare l'età, il periodo e la coorte come variabili continue e utilizzare metodi di *smoothing* per adattare questi dati. Questo metodo appartiene alla classe dei modelli additivi generalizzati (*Generalised Additive Models*, GAM). In particolare in questo metodo, Carstensen ha proposto di usare le splines naturali per smussare le curve non lineari del modello età-periodo-coorte [16, 27].

La forma generale di un modello moltiplicativo età-periodo-coorte per i tassi  $\lambda(a, p)$  all'età  $a$  nel periodo  $p$  per gli individui nella coorte di nascita  $c=p-a$  è:

$$\log[\lambda(a, p)] = f(a) + g(p) + h(c)$$

per tre funzioni ( $f$ ,  $g$  e  $h$ ).

Si assume che  $a$ ,  $p$  e  $c$  rappresentino la media dell'età, del periodo e della coorte per le unità di osservazione. Il modello permette che gli effetti delle tre variabili siano non lineari quando il tasso è espresso su scala logaritmica. A questo punto diventa irrilevante quale forma parametrica sia scelta per le funzioni  $f$ ,  $g$  e  $h$ .

Il modello predice i tassi in ogni punto nel diagramma di Lexis. Per i dati derivanti da tabelle, si deve assumere che il tasso sia costante all'interno di ogni sottogruppo



della tabella facente parte di un diagramma di Lexis. Si ricorda che il logaritmo del tasso segue una distribuzione di Poisson.

Il problema pratico è come scegliere una parametrizzazione ragionevole di queste funzioni e come ottenere le stime.

Dal momento che sappiamo che  $c = p - a$  e quindi che  $p - a - c = 0$ , possiamo riscrivere il modello nel seguente modo:

$$\begin{aligned} \log[\lambda(a, p)] &= f(a) + g(p) + h(c) = f(a) + g(p) + h(c) + \gamma(p - a - c) \\ &= f(a) - \mu_a - \gamma a + g(p) + \mu_a + \mu_c + \gamma p + h(c) - \mu_c - \gamma c \end{aligned}$$

Poiché le tre variabili età, periodo e coorte in origine sono variabili continue, si possono modellare i loro effetti con funzioni parametriche *smooth* della classe delle medie, per esempio le splines naturali, lineari, quadratiche, cubiche e i polinomi frazionari. In questo modello consideriamo che  $f$ ,  $g$  e  $h$  siano delle funzioni splines naturali non lineari rispettivamente per gli effetti dell'età, del periodo e della coorte.

Quindi:

$$\log[\lambda(a, p)] = \check{f}(a) + \check{g}(p) + \check{h}(c)$$

dove

$$\begin{aligned} \check{f}(a) &= f(a) - \mu_a - \gamma a \\ \check{g}(p) &= g(p) + \mu_a + \mu_c + \gamma p \\ \check{h}(c) &= h(c) - \mu_c - \gamma c \end{aligned}$$

Fissiamo  $\mu_a$ ,  $\mu_c$  e  $\gamma$  in modo che le funzioni conseguenti  $\check{f}(a)$ ,  $\check{g}(p)$  e  $\check{h}(c)$  soddisfino i vincoli richiesti.

Ora procediamo con la seguente parametrizzazione:

1. otteniamo le matrici dei parametri dei modelli dell'età  $M_a$ , periodo  $M_p$  e coorte  $M_c$ , ciascuna con l'intercetta;
2. estraiamo il trend lineare (cioè sottraiamo l'effetto deriva) da  $M_p$  e  $M_c$ , proiettando le loro colonne sul complemento ortogonale di  $[1|p]$  e  $[1|c]$  e otteniamo così  $\tilde{M}_p$  e  $\tilde{M}_c$ ;

3. centriamo l'effetto della coorte in  $c_0$ : prendiamo una riga da  $\tilde{M}_c$  corrispondente a  $c_0$ , creo una matrice con  $c_0$  di dimensioni pari a  $\tilde{M}_c$  e la sottraggo da  $\tilde{M}_c$  per formare  $\tilde{M}_{c_0}$ ;
4. usiamo:  $M_a$  per gli effetti dell'età,  $\tilde{M}_p$  per gli effetti del periodo e  $[c - c_0 | \tilde{M}_{c_0}]$  per gli effetti della coorte;
5. il valore di  $\hat{f}(a)$  è  $M_a \hat{\beta}_a$ , e si procede in modo analogo per gli altri due effetti. La varianza è data da  $M'_a \hat{\Sigma}_a M_a$ , dove  $\hat{\Sigma}_a$  è la matrice di varianze e covarianze di  $\hat{\beta}_a$ .

Per questo metodo dell'analisi APC, gli intervalli di confidenza vengono calcolati usando il metodo della verosimiglianza profilo. Gli intervalli di confidenza calcolati con il metodo della verosimiglianza profilo non richiedono che gli stimatori siano distribuiti normalmente, e sembrano più adeguati per campioni con una numerosità ridotta rispetto agli intervalli calcolati con il metodo classico (Wald-type). Si basano su una approssimazione asintotica della distribuzione Chi-quadro della statistica test del rapporto del logaritmo delle verosimiglianze. La funzione di verosimiglianza profilo si ottiene massimizzando la funzione di verosimiglianza rispetto al parametro di interesse. Partendo dalla statistica test del rapporto delle verosimiglianze, con alcuni passaggi matematici, si ottengono gli intervalli di confidenza con il metodo della verosimiglianza profilo per il parametro di interesse. Gli intervalli di confidenza calcolati con questo metodo non variano con la trasformazione monotona di un parametro.

### 3.6. Metodo dei minimi quadrati parziali

Un altro metodo di decomposizione degli effetti di età, periodo e coorte si basa sui modelli fattoriali, come ad esempio il modello delle componenti principali e quello dei minimi quadrati parziali. In questo paragrafo descriveremo il modello dei minimi quadrati parziali (*Partial Least Squares*, PLS).

Il modello APC può anche essere scritto come:

$$y_{ij} = \alpha + A_i + P_j + C_k + \varepsilon_{ij} \quad (3)$$

dove  $\alpha$  è il termine costante,  $A_i$  l'effetto dell'età,  $P_j$  l'effetto del periodo,  $C_k$  l'effetto della coorte di nascita e  $\varepsilon_{ij}$  un termine di disturbo con distribuzione Normale  $(0, \sigma^2)$ . Si consideri il modello (3) come un modello di coorte e si assuma che l'ampiezza dei gruppi di età coincida con quella dei periodi. Inoltre i parametri  $A_i$ ,  $P_j$  e  $C_k$  devono essere sottoposti al seguente vincolo:

$$\sum_{i=1}^I A_i = \sum_{j=1}^J P_j = \sum_{k=1}^K C_k = 0$$

Se applichiamo questo vincolo al modello (3), possiamo riscrivere il modello usando la notazione matriciale come:

$$y = X\beta + \varepsilon \quad (4)$$

dove  $y$  è un vettore di dimensioni  $(IJ*1)$ ,  $X$  è una matrice  $IJ*[2(I+J)-3]$ ,  $\beta$  è un vettore  $[2(I+J)-3]*1$  specificato come  $\beta = (\alpha, A_1, \dots, A_{I-1}, P_1, \dots, P_{J-1}, C_1, \dots, C_{K-1})$ , e  $\varepsilon$  è un vettore  $IJ*1$ . Solitamente per stimare  $\beta$  si usa il metodo della massima verosimiglianza, ma in questo caso non si otterrebbero delle stime uniche a causa della dipendenza lineare dei parametri degli effetti.

In questo caso per risolvere il problema dell'identificabilità usiamo la regressione dei minimi quadrati parziali con una procedura a due stadi [24, 28, 29]. Nel primo stadio, il metodo fattoriale è applicato ai dati di  $y$  e  $X$  per ottenere i fattori non osservati. Successivamente, si ottengono automaticamente le stime di  $\beta$ , poiché ogni fattore non osservato è rappresentato come una combinazione lineare delle componenti osservate di  $X$ ; in sostanza ogni fattore viene usato come regressore. Nella regressione PLS ogni fattore è dato da una combinazione lineare delle variabili esplicative d'origine, ma, a differenza della regressione delle componenti principali, è ottenuto considerando la relazione tra le variabili dipendenti e indipendenti. I modelli fattoriali sono usati per risolvere il problema di molte variabili osservate con

possibile collinearità; nel caso del problema di identificabilità del modello di coorte, si considera il modello (4) con  $N=IJ$  e  $M=2(I+J)-3$ .

Prima di applicare la regressione PLS, si assume che tutte le variabili siano normalizzate con media zero e che ci siano  $k$  fattori PLS.

Quindi possiamo riscrivere il modello come:

$$y_n = \beta' x_n + \varepsilon_n \quad n = 1, \dots, N$$

dove  $\beta = (\beta_1 \dots \beta_M)'$  e  $x_n = (x_{1n} \dots x_{Mn})'$ .

Procediamo secondo l'algoritmo creato da Helland:

Passo 1:

poniamo  $u_n = y_n$  e  $v_{in} = x_{in}$ ,  $i = 1, \dots, M$  e  $j = 1$ .

Passo 2:

otteniamo il  $j$ -esimo fattore PLS come  $f_{jn} = w_j' v_n$ , dove  $w_j = (w_{1j} \dots w_{Mj})'$ ,  $w_{ij} = \text{Cov}(u_n, v_{in})$ ,  $i = 1, \dots, M$ .

Passo 3:

regrediamo  $u_n$  e  $v_{in}$ ,  $i = 1, \dots, M$ , su  $f_{jn}$ . Chiamiamo i residui di queste regressioni rispettivamente  $\tilde{u}_n$  e  $\tilde{v}_{in}$ .

Passo 4:

se  $j = k$  ci fermiamo, altrimenti poniamo  $u_n = \tilde{u}_n$ ,  $v_{in} = \tilde{v}_{in}$ ,  $i = 1, \dots, M$  e  $j = j + 1$  e torniamo al Passo 2.

Al Passo 3, per  $j = k$  otteniamo la seguente equazione:

$$y_n = \{\hat{\phi}_1 w_1 + \hat{\phi}_2 w_2 (I - \hat{\theta}_1 w_1) + \dots + \hat{\phi}_k w_k (I - \hat{\theta}_1 w_1) \dots (I - \hat{\theta}_{k-1} w_{k-1})\} x_n + \tilde{u}_n,$$

dove  $\hat{\phi}_j$  è il parametro ottenuto dalla regressione di  $u_n$  su  $f_{jn}$  e  $\hat{\theta}_j$  è il vettore di parametri ottenuto dalla regressione di  $v_{in}$  su  $f_{jn}$ .

Il valore di  $k$  non può essere definito con una regola statistica e il numero di iterazioni necessarie per ottenere la convergenza in questo algoritmo non può essere predeterminato. Quindi consideriamo arbitrariamente che  $k$  sia uguale a  $M/4$ .

Il modello PLS seleziona le componenti  $f_{jn}$  che massimizzano la covarianza tra l'outcome  $y_n$  e  $f_{jn}$  :

$$\max[cov(f_{jn}, y_n)]^2 = \max[corr(w'_j v_n, y_n)]^2 var(w'_j v_n)$$

sotto i vincoli che la somma dei pesi al quadrato sia pari ad uno,  $\sum w_{ij}^2 = 1$  (in altri termini  $\|w_i\| = 1$ ) e che la correlazione tra tutte le componenti sia pari a zero ( $w_i \perp w_j (i \neq j)$  ).

Le componenti PLS sono ordinabili secondo la grandezza della varianza che spiegano nell'outcome. Nel PLS, la prima componente spiega la maggior parte della varianza dell'outcome.

Avendo così ottenuto i coefficienti per le componenti PLS, dobbiamo ricondurci ai coefficienti delle variabili predittive di origine. Il coefficiente della regressione PLS per ogni  $x_i$  è derivato dalla somma dei prodotti dei coefficienti della regressione per le componenti PLS e i pesi per ogni  $x_i$ . Per esempio, quando l'outcome è regresso su due componenti PLS, l'equazione è data da:

$$\begin{aligned} y &= \beta_1 f_1 + \beta_2 f_2 + \varepsilon \\ &= \beta_1 (w_{11} Age + w_{12} Period + w_{13} Cohort) + \beta_2 (w_{21} Age + w_{22} Period + \\ &\quad w_{23} Cohort) + \varepsilon \\ &= (\beta_1 w_{11} + \beta_2 w_{21}) Age + (\beta_1 w_{12} + \beta_2 w_{22}) Period + \\ &\quad (\beta_1 w_{13} + \beta_2 w_{23}) Cohort + \varepsilon \end{aligned}$$

dove  $\beta_1$  e  $\beta_2$  sono rispettivamente i coefficienti stimati per la prima e la seconda componente PLS e  $\varepsilon$  è il termine residuo. Quindi il coefficiente di regressione per l'età, per esempio, è  $\beta_1 w_{11} + \beta_2 w_{21}$ .

Anche per questo metodo dell'analisi APC non è stato raggiunto un consenso su come calcolare gli intervalli di confidenza. Pertanto per stimare gli errori standard e gli intervalli di confidenza direttamente dai dati si è ricorso al metodo jackknife [30]. Il metodo jackknife, simile al bootstrap, è una procedura di ricampionamento utilizzata nell'inferenza statistica per stimare l'errore standard di una grandezza.

L'idea fondamentale dietro il metodo jackknife sta nel ricalcolare più volte la grandezza statistica stimata lasciando fuori dal campione una osservazione alla volta. Il ricalcolo si effettua a partire da questo nuovo insieme (il campione privato di una osservazione).

# CAPITOLO 4

## Analisi predittiva

---

### 4.1. Introduzione

Come abbiamo già detto precedentemente, prevedere gli andamenti futuri dei tassi di incidenza e di mortalità è necessario per acquisire informazioni utili all'epidemiologia e alla programmazione sanitaria.

La modellazione statistica degli andamenti passati permette di stimare le proiezioni dei trend di incidenza e di mortalità estrapolando gli andamenti temporali dai tassi osservati. Il numero di nuovi casi di tumori o di morti è calcolato applicando i tassi stimati alle popolazioni predette. Le proiezioni basate sull'estrapolazione degli andamenti di incidenza e mortalità per tumore nel tempo presuppongono che i trend rimangano stabili nel comportamento dei rischi, nessun programma di intervento o di screening venga avviato, e non vi sia alcun cambiamento nelle tecniche diagnostiche [31]. I trend d'incidenza o mortalità per tumore possono essere descritti come trend sull'età alla diagnosi o di morte, anno di diagnosi o di morte (periodo), e/o anno di nascita (coorte). L'età è la scala temporale più importante nello studio del rischio di cancro; caratterizza l'esposizione cumulativa del corpo ad agenti cancerogeni nel tempo. Gli effetti del periodo corrispondono ad eventi che modificano il rischio d'incidenza o di mortalità, indipendentemente dalla fascia di età e che sono, di solito, dovuti ad un ambiente che cambia. Gli effetti di coorte coinvolgono i fattori di rischio che sono condivisi da una generazione specifica. Va osservato che l'assunzione di tendenze invariate nell'andamento dei tassi è molto forte e può non essere realistica [31]. L'andamento dei tassi osservati riflette l'andamento non osservato del rischio di tumore. Di solito, il trend può essere classificato come (totale o specifico per età) trend di periodo e / o trend di coorte, portando a due classi di modelli: modelli età-periodo e modelli età-periodo-coorte. Matematicamente, l'andamento può essere descritto come lineare o non lineare e diverse tecniche di

modellazione statistica possono essere utilizzate, tra cui modelli parametrici, semi-parametrici e non-parametrici. Poiché metodi statistici differenti possono causare diverse proiezioni di tumore, può essere difficile determinare quale metodo sia più appropriato. La scelta del modello statistico appropriato è fondamentale per produrre proiezioni valide [31].

In generale, gli effetti di periodo possono modificare il rischio di tumore sia nel breve che nel lungo periodo, effetti di coorte sul rischio di tumore sono più importanti per un lungo periodo che per un breve periodo. Così in generale le proiezioni a breve termine sono basate su modelli età-periodo, mentre le proiezioni a lungo termine tengono conto degli effetti di coorte sul lungo periodo e si basano su modelli età-periodo-coorte.

Negli ultimi decenni sono stati introdotti vari metodi di proiezione della mortalità per tumore. Tra questi, per il breve periodo, troviamo: il metodo della regressione di Poisson, quelli basati su modelli ARIMA (*AutoRegressive Integrated Moving Average*) per serie temporali e il joinpoint. Mentre per il lungo periodo troviamo metodi che si basano sul modello età-periodo-coorte, come il Nordpred [31-34].

Nei seguenti paragrafi verranno presentati e confrontati tra loro alcune tecniche di proiezione della mortalità di breve periodo, che utilizzano un modello età-periodo con joinpoint o un modello età-periodo bayesiano, e una tecnica di lungo periodo, che utilizza un modello età-periodo-coorte bayesiano.

## **4.2. Modello età-periodo con Joinpoint**

### **4.2.1. Introduzione alla Joinpoint**

Un modo per descrivere gli andamenti dei tassi di mortalità nel tempo consiste nel calcolare il cambiamento percentuale annuo (*Estimated Annual Percent Change*, EAPC). Si assume che i tassi presentino un cambiamento costante da un anno all'altro. Tassi che cambiano secondo una costante percentuale ogni anno, cambiano linearmente su una scala logaritmica. Per questa ragione, per stimare l'EAPC di una serie di dati, viene utilizzato il seguente modello di regressione:



$$\log(\lambda_y) = b_0 + b_1 y$$

dove  $\log(\lambda_y)$  è il logaritmo naturale del tasso nell'anno  $y$ .

L'EAPC tra l'anno  $y$  e l'anno  $y+1$  è pari a :

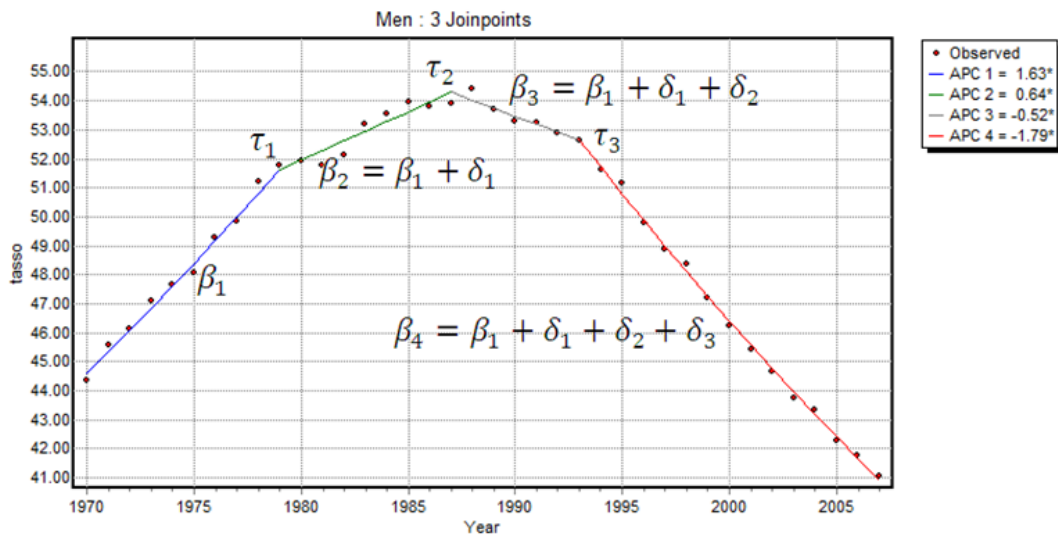
$$EAPC_{y,y+1} = \left[ \frac{(\lambda_{y+1} - \lambda_y)}{\lambda_y} \right] \times 100 = (e^{b_1} - 1) \times 100$$

Non è ragionevole attendersi che un singolo EAPC possa descrivere dettagliatamente l'andamento di una intera serie di dati. Per questo, è molto utile applicare un modello joinpoint che attraverso criteri statistici determina quando e quante volte l'EAPC cambia significativamente.

Per i tassi di mortalità, il modello viene stimato attraverso segmenti log-lineari uniti tra loro. Quindi, stimare il modello joinpoint che meglio descrive i dati, ci permette di individuare e determinare per quanto tempo l'EAPC rimane costante e quando presenta un cambiamento.

Per identificare gli anni in cui si verificano cambiamenti statisticamente significativi dei tassi di mortalità nel periodo di osservazione in studio, si è utilizzato un modello di regressione joinpoint proposto da Kim HJ et al [9]. La regressione joinpoint è stata implementata nel software Joinpoint del National Cancer Institute.

In breve, il modello di regressione impiegato assume la linearità dei trend, più precisamente del logaritmo dei tassi. Ogni curva può essere approssimata localmente in modo soddisfacente mediante un segmento lineare purché di lunghezza appropriata. Il modello impiegato, infatti, si basa sull'individuazione dei segmenti lineari che meglio si adattano ai tassi osservati, rendendo minima la somma dei quadrati delle distanze dei punti dai segmenti stessi [35]. I punti in cui il tasso cambia in modo statisticamente significativo si chiamano "joinpoints". Il massimo numero di segmenti in cui il trend risulta scomposto è limitato dal numero di joinpoints arbitrariamente fissato prima dell'analisi. L'anno in cui si stima verificarsi il joinpoint è l'anno che individua una variazione del trend.



**Figura 4:** Esempio di modello joinpoint.

Il modello di regressione joinpoint per le coppie di osservazione  $(x_i, y_i)$ , per  $i=1, \dots, n$ , dove le  $y_i$  rappresentano i tassi di mortalità osservati al tempo  $x_i$ , può essere scritto nel seguente modo:

$$E[y|x] = \beta_0 + \beta_1 x + \sum_{i=1}^k \delta_i (x - \tau_i)^+$$

dove:

$k$  è il numero di joinpoints non noti

i  $\tau_k$  rappresentano i joinpoints non noti

$$(x - \tau_i)^+ = \begin{cases} x - \tau_k & \text{per } x > \tau_k \\ 0 & \text{altrimenti} \end{cases}$$

Nel caso di un modello privo di joinpoint l'equazione si riduce all'equazione di una retta con intercetta  $\beta_0$  e pendenza  $\beta_1$ . I termini  $\delta_i(x - \tau_i)^+$  rappresentano, invece, la variazione della pendenza della retta per gli eventuali successivi segmenti e sono uguali a zero negli anni antecedenti quello di joinpoint. Il modello migliore è individuato mediante un algoritmo con un test di permutazione, che prevede il confronto tra modelli con un numero diverso di joinpoints. Il primo confronto è effettuato tra il modello privo di joinpoints e quello con un numero di joinpoints pari al numero massimo fissato. Il modello finale è quello con il minor numero di

parametri, per il quale l'aggiunta di un ulteriore parametro (punto di joinpoint) non comporti miglioramenti significativi ( $\alpha = 0.05$  aggiustata per test multipli). Una volta selezionato il modello migliore la misura dell'EAPC stimato per ogni segmento è pari a:

$$EAPC = (e^{b_0} - 1) \times 100$$

Nell'esempio applicativo di questa tesi si è applicato il modello di regressione joinpoint sul numero di morti per il periodo 1970-1999, con i seguenti vincoli:

- il numero di anni di osservazione successivi all'ultimo joinpoint deve essere almeno pari a cinque;
- il numero di anni di osservazione tra due joinpoints deve essere almeno quattro;
- il numero massimo di joinpoints è stato fissato, prima dell'analisi, pari a cinque.

#### **4.2.2. Joinpoint sul numero di morti**

Le analisi dei trend effettuate con il modello di regressione joinpoint possono essere utilizzate per predire i dati di incidenza e mortalità.

Questo metodo di previsione procede nel seguente modo. Si adatta un modello di regressione joinpoint al logaritmo del numero di decessi età-specifici per classi di età di 5 anni per individuare i segmenti dei trend più recenti. In seguito, un modello di regressione viene applicato ai dati di mortalità per ogni gruppo di età nel periodo di tempo dell'ultimo segmento individuato dal modello joinpoint, al fine di stimarne i coefficienti di regressione. Si applica lo stesso modello con i coefficienti stimati agli anni per cui si vuole predire la mortalità, al fine di calcolare il numero di morti età-specifico previsto, i corrispondenti intervalli di confidenza al 95% (IC) e intervalli di previsione (IP, cioè gli intervalli di confidenza per la previsione di ogni singolo valore futuro, calcolato con un errore standard che tiene conto della variabilità della nuova osservazione) [36]. Con il numero di decessi previsti specifici per età si calcolano i tassi di mortalità specifici per età e standardizzati per età con i corrispondenti intervalli di confidenza e di previsione al 95%, usando i dati di popolazione proiettati per il periodo di interesse.

Si sono considerati tre diversi modelli di regressione (lineare, Poisson e log-lineare) per fare le proiezioni dei dati di mortalità. Nel seguito, mostreremo brevemente alcune loro caratteristiche.

- Modello di regressione lineare

Si assume che il numero di decessi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuisca e si proietti nel futuro secondo un modello di regressione lineare:

$$E[d] = b_0 + b_1 y$$

dove:  $E[.]$  è il valore atteso;

$d$  è il numero di decessi;

$y$  è l'anno di calendario.

I coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo dei minimi quadrati dal modello sui dati di mortalità per il periodo di tempo dell'ultimo segmento individuato dal modello joinpoint.

Per calcolare i limiti degli intervalli di confidenza per il numero di morti previsto si utilizza la seguente formula:

$$d^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

dove:  $d^*$  è il numero di decessi previsto;

$t_{1-\frac{\alpha}{2}, n-2}$  è il valore della t-Student per  $(1 - \frac{\alpha}{2})$  con  $(n-2)$  gradi di libertà;

$n$  è il numero totale delle osservazioni;

$y$  sono gli anni di osservazione;

$y^*$  è l'anno di previsione;

$\bar{y}$  è la media delle osservazioni;

$s$  è la radice dello scarto quadratico medio del modello data da:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2}$$

dove:  $d$  sono i decessi;

$\bar{d}$  è la media dei decessi.

Mentre gli intervalli di previsione sono dati da:

$$d^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{1 + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

#### - Modello di regressione di Poisson

Dal momento che gli eventi osservati sono approssimativamente distribuiti secondo la legge di probabilità di Poisson, si assume che il numero di decessi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuisca e si proietti nel futuro secondo un modello di regressione di Poisson nel seguente modo:

$$\log(E[d]) = b_0 + b_1 y$$

dove:  $E[.]$  è il valore atteso;

$d$  è il numero di decessi;

$y$  è l'anno di calendario.

In questo caso, i coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo della massima verosimiglianza dal modello sui dati di mortalità di interesse individuati dal modello joinpoint.

Ai fini di stimare la variabilità di un coefficiente di regressione stimato attraverso un modello di Poisson, in assenza di una formula esatta per calcolare gli intervalli di confidenza, si è ricorso al metodo della verosimiglianza profilo. Gli intervalli di confidenza calcolati con il metodo della verosimiglianza profilo non richiedono che gli stimatori siano distribuiti normalmente, e sembrano più adeguati per campioni con una numerosità ridotta rispetto agli intervalli calcolati con il metodo classico (Wald-type). Si basano su una approssimazione asintotica della distribuzione Chi-

quadro della statistica test del rapporto del logaritmo delle verosimiglianze. La funzione di verosimiglianza profilo si ottiene massimizzando la funzione di verosimiglianza rispetto al parametro di interesse. Partendo dalla statistica test del rapporto delle verosimiglianze, con alcuni passaggi matematici, si ottengono gli intervalli di confidenza con il metodo della verosimiglianza profilo per il parametro di interesse. Gli intervalli di confidenza calcolati con questo metodo non variano con la trasformazione monotona di un parametro.

Per stimare gli intervalli di previsione del numero previsto di decessi si ricorre al metodo non-parametrico bootstrap [26]. Il metodo consiste nel ricampionare 1000 volte i dati attraverso un campionamento casuale semplice, al fine di ottenere le previsioni del numero di decessi per gli anni di interesse attraverso il modello di regressione di Poisson. L'intervallo di previsione per ognuno di questi anni è poi stimato prendendo il quantile 2.5 e il quantile 97.5 delle 1000 stime bootstrap.

- Modello di regressione log-lineare

Si assume che il numero di decessi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuisca e si proietti nel futuro secondo un modello di regressione log-lineare nel seguente modo:

$$\log(E[d]) = b_0 + b_1 y$$

dove:  $E[.]$  è il valore atteso;

$d$  è il numero di decessi;

$y$  è l'anno di calendario.

I coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo dei minimi quadrati dal modello sui dati di mortalità individuati dal modello joinpoint.

Per calcolare gli intervalli di confidenza per il numero di morti previsto si utilizza la seguente formula:

$$\exp \left( d^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \right)$$

con la stessa notazione usata per il modello lineare.

Mentre per gli intervalli di previsione:

$$\exp \left( d^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{1 + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \right)$$

### 4.2.3. Joinpoint sui tassi

Questo altro metodo di proiezione dei dati mortalità è simile al precedente, ma adatta un modello di regressione joinpoint al logaritmo dei tassi età-specifici per classi di età di 5 anni per individuare i segmenti più recenti dei trend. Come per il modello sul numero di morti, un modello di regressione viene applicato ai tassi di mortalità per ogni gruppo di età nel periodo di tempo dell'ultimo segmento individuato dal modello joinpoint, al fine di stimarne i coefficienti di regressione. Si applica lo stesso modello con i coefficienti appena stimati agli anni per cui si vuole predire la mortalità, al fine di calcolare i tassi età-specifici previsti, i corrispondenti intervalli di confidenza e di previsione al 95%. Con i tassi di mortalità previsti specifici per età si calcolano il numero di decessi specifici per età e il tassi standardizzati per età con i corrispondenti intervalli di confidenza e di previsione al 95%, usando i dati di popolazione previsti per il periodo di interesse.

Anche per questo approccio si sono considerati tre diversi modelli di regressione (lineare, Poisson e log-lineare) per fare le proiezioni dei dati di mortalità. Nel seguito, mostreremo brevemente alcune loro caratteristiche.

#### - Modello di regressione lineare

Si assume che i tassi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuiscano e si proiettino nel futuro secondo un modello di regressione lineare:

$$E[\lambda] = b_0 + b_1 y$$

dove:  $\lambda$  è il tasso di mortalità;

$y$  è l'anno di calendario.

I coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo dei minimi quadrati. Per calcolare gli intervalli di confidenza per i tassi previsti si utilizza la seguente formula:

$$\lambda^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

dove:  $\lambda^*$  è il tasso di mortalità previsto;

$t_{1-\frac{\alpha}{2}, n-2}$  è il valore della t-Student per  $(1 - \frac{\alpha}{2})$  con  $(n-2)$  gradi di libertà;

$n$  è il numero totale delle osservazioni;

$y$  sono gli anni di osservazione;

$y^*$  è l'anno di previsione;

$\bar{y}$  è la media delle osservazioni;

$s$  è la radice dello scarto quadratico medio del modello data da:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\lambda_i - \bar{\lambda})^2}$$

dove:  $\lambda$  è il tasso di mortalità;

$\bar{\lambda}$  è la media dei tassi.

Mentre per gli intervalli di previsione si ha:

$$\lambda^* \pm t_{1-\frac{\alpha}{2}, n-2} s \sqrt{1 + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

dove la notazione è la stessa della formula precedente.

#### - Modello di regressione di Poisson

Poiché i tassi sono approssimativamente distribuiti secondo la legge di probabilità di Poisson, così come gli eventi osservati, si assume che i tassi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuiscano e si distribuiranno nel futuro secondo un modello di regressione di Poisson. Si ha come variabile dipendente la variabile  $Y = \log(E[\lambda])$ , pari alla trasformazione logaritmica del tasso atteso  $\lambda$ .



$$\begin{aligned}\log(E[\lambda]) &= b_0 + b_1 y \\ \log d &= b_0 + b_1 y + \log m\end{aligned}$$

dove:  $\lambda$  è il tasso di mortalità;  
 $y$  è l'anno di calendario;  
 $d$  è il numero di decessi;  
 $m$  sono le persone-tempo.

Possiamo notare che nel modello il logaritmo del tasso viene scisso in logaritmo del numero di morti e logaritmo della popolazione, dove quest'ultimo prende il nome di *offset*.

In questo caso, i coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo della massima verosimiglianza dal modello sui dati di mortalità di interesse individuati dal modello joinpoint.

Similmente al modello con joinpoint sul numero di decessi si usano i metodi della verosimiglianza profilo e bootstrap per calcolare gli intervalli di confidenza e di previsione.

#### - Modello di regressione log-lineare

Si assume che i tassi dell'ultimo segmento del trend individuato dal modello joinpoint si distribuiscano e si proiettino nel futuro secondo un modello di regressione log-lineare nel seguente modo:

$$\log(E[\lambda]) = b_0 + b_1 y$$

dove:  $\lambda$  è il tasso di mortalità;  
 $y$  è l'anno di calendario.

I coefficienti di regressione  $b_0$  e  $b_1$  sono stimati con il metodo dei minimi quadrati dal modello sui dati di mortalità individuati dal modello joinpoint.

Per calcolare gli intervalli di confidenza per i tassi previsti si utilizza la seguente formula:

$$\exp\left(\lambda^* \pm t_{1-\frac{\alpha}{2}, n-2} S \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}\right)$$

con la stessa notazione usata per il modello lineare.

Mentre per gli intervalli di previsione:

$$\exp\left(\lambda^* \pm t_{1-\frac{\alpha}{2}, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}\right)$$

In questa tesi, le analisi dei diversi modelli età-periodo con joinpoint sono state svolte con i software statistici R e SAS.

### 4.3. Modello età-periodo-coorte bayesiano

Un altro metodo di previsione dei dati di mortalità (o incidenza) è quello proposto da Isabelle Bray [37], che si basa su un modello di analisi età-periodo-coorte con approccio bayesiano. Bray, ispirandosi al modello di previsione di Berzuini e Clayton [38], ha sviluppato un approccio bayesiano per smussare gli effetti di età, periodo e coorte e stimare i tassi specifici per età dalla loro distribuzione a posteriori attraverso la tecnica Markov Chain Monte Carlo (MCMC).

Si considerano dati di mortalità (o incidenza) stratificati per classi di età quinquennali e periodi quinquennali con i corrispondenti anni-persona a rischio (popolazioni). Le classi di età sono indicizzate come  $a=1, \dots, A$ , i periodi come  $p=1, \dots, P$  e le coorti di nascita come  $c=1, \dots, C$  (dove  $C=A+P-1$ ). Per ognuna classe di età  $a$  e per ogni periodo  $p$ , si ricorda che il conteggio del numero di morti (o casi)  $d_i$  si distribuisce come:

$$d_i \sim \text{Poisson}(\mu_i) \quad \text{con } i = 1, \dots, AP$$

Questo metodo si basa su un modello età-periodo-coorte log-lineare definito come:

$$\log \mu_i = \log m_i + \alpha_{età_i} + \beta_{periodo_i} + \gamma_{coorte_i}$$

o in modo equivalente:

$$\log \mu_{ap} = \log m_{ap} + \alpha_a + \beta_p + \gamma_c$$

in cui il logaritmo della media  $\mu_i$  di  $d_i$  è regresso sugli effetti di età ( $\alpha_{età_i}$ ), periodo ( $\beta_{periodo_i}$ ) e coorte ( $\gamma_{coorte_i}$ ) usando il logaritmo della corrispondente popolazione  $m_i$  come *offset*.

Come abbiamo detto precedentemente, il modello età-periodo-coorte ha un problema strutturale intrinseco di identificabilità. Questo non permette di ottenere soluzioni uniche delle stime degli effetti età, periodo e coorte. Tuttavia, fortunatamente, le proiezioni basate su modelli età-periodo-coorte non risentono di questo problema e sono determinate in modo univoco [39]. Infatti, le stime degli attesi sia predetti che proiettati sono univoche, in quanto le combinazioni con stime diverse degli effetti di età, periodo e coorte che formano le infinite soluzioni del modello sommano allo stesso risultato.

I metodi bayesiani tengono conto delle conoscenze a priori sui parametri. Le stime basate su simulazioni bayesiane consentono di specificare priori flessibili che smussano gli effetti del tempo. Bray specifica un modello a priori autoregressivo per smussare gli effetti di età, periodo e coorte, in quanto ci si aspettano cambiamenti gradualmente e continui sulle scale temporali esaminate e non cambiamenti repentini o salti discreti. In seguito il modello predice le funzioni di distribuzione di probabilità a posteriori degli effetti di età, periodo e coorte, e di conseguenza dei tassi, facendo inferenza bayesiana con tecniche computazionali MCMC.

Spieghiamo ora più dettagliatamente come procede questo metodo di previsione.

Come consigliato nella letteratura da Breslow e Clayton [40] e da Berzuini e Clayton [38], questo metodo specifica un modello a priori autoregressivo Gaussiano per smussare gli effetti su ciascuna scala temporale e estrapolare gli effetti di periodo e coorte. Secondo questo modello ogni punto (eccetto i primi due di ogni scala) è predetto con un'estrapolazione lineare dei suoi due immediati predecessori e con un errore casuale distribuito come una Normale con media uguale a zero. La precisione

di questa distribuzione Normale può essere rappresentata come un iperparametro, la cui distribuzione a priori (detta *iperpriori*) rispecchia la conoscenza a priori riguardante lo *smoothness* dei parametri. Nel nostro caso useremo distribuzioni *iperpriori* non informative, affinché gli iperparametri siano stimati solamente a partire dai dati.

Nel modello autoregressivo per i primi due parametri  $\alpha_1$  e  $\alpha_2$  per l'effetto età,  $\beta_1$  e  $\beta_2$  per l'effetto periodo e  $\gamma_1$  e  $\gamma_2$  per l'effetto coorte vengono assegnate delle distribuzioni a priori non informative. Supponendo che siano richieste le proiezioni per  $N$  periodi futuri, le distribuzioni a priori per i parametri del modello possono essere definite come segue.

Per gli  $A$  effetti di età:

$$\alpha_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\alpha}\right)$$

$$\alpha_2 | \alpha_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\alpha}\right)$$

$$\alpha_a | \alpha_{1,\dots,a-1} \sim \text{Normale}\left(2\alpha_{a-1} - \alpha_{a-2}, \frac{1}{\tau_\alpha}\right) \quad \text{per } 3 \leq a \leq A$$

Per i  $P+N$  effetti di periodo:

$$\beta_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\beta}\right)$$

$$\beta_2 | \beta_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\beta}\right)$$

$$\beta_p | \beta_{1,\dots,p-1} \sim \text{Normale}\left(2\beta_{p-1} - \beta_{p-2}, \frac{1}{\tau_\beta}\right) \quad \text{per } 3 \leq p \leq P + N$$

Per i  $C+N$  effetti di coorte:

$$\gamma_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\gamma}\right)$$

$$\gamma_2 | \gamma_1 \sim \text{Normale} \left( 0, 1000000 \frac{1}{\tau_\gamma} \right)$$

$$\gamma_c | \gamma_{1,\dots,c-1} \sim \text{Normale} \left( 2\gamma_{c-1} - \gamma_{c-2}, \frac{1}{\tau_\gamma} \right) \quad \text{per } 3 \leq c \leq C + N$$

Agli iperparametri di precisione  $\tau_\alpha$ ,  $\tau_\beta$  e  $\tau_\gamma$  sono state assegnate delle distribuzioni iperpriori vaghe, ossia:

$$\tau_\alpha \sim \text{Gamma}(0.001, 0.001)$$

$$\tau_\beta \sim \text{Gamma}(0.001, 0.001)$$

$$\tau_\gamma \sim \text{Gamma}(0.001, 0.001)$$

Queste iperpriori sono vaghe in quanto il loro logaritmo approssima una distribuzione univariata.

A causa della non identificabilità dei parametri, in questo modello si sono imposti dei vincoli nell'algoritmo MCMC. Si pongono come vincoli che il trend lineare degli effetti età sia pari a zero e che le somme degli effetti di periodo e di coorte per i dati osservati sottratti delle rispettive medie siano pari a zero.

EsPLICITIAMO brevemente questi vincoli con le seguente formule [13].

Per quanto riguarda gli effetti dell'età ( $\sum_a \alpha_a = 0$ ), il trend lineare dell'effetto età  $\alpha_a$  può essere descritto dal seguente contrasto:

$$\alpha_L = D \sum_a d_a \alpha_a$$

$$\text{dove: } d_a = a - \frac{1}{2}A - \frac{1}{2} = a - \frac{A+1}{2}$$

$$D = (\sum_a d_a^2)^{-1} = \left( \frac{A(A+1)(A-1)}{12} \right)^{-1}$$

La componente dagli effetti età separata dal trend lineare è:

$$\tilde{\alpha}_a = \alpha_a - d_a \alpha_L$$

Invece, per gli effetti dei periodi e delle coorti si vuole che  $\sum_b \beta_b = 0$  e che  $\sum_c \gamma_c = 0$ , ottenibili attraverso i seguenti vincoli:

$$\tilde{\beta}_b = \beta_b - \bar{\beta}_b$$

$$\tilde{\gamma}_c = \gamma_c - \bar{\gamma}_c$$

Infine, ricombinando gli effetti “smussati” di età, periodo e coorte, si ottengono le stime e le proiezioni dei tassi nel seguente modo:

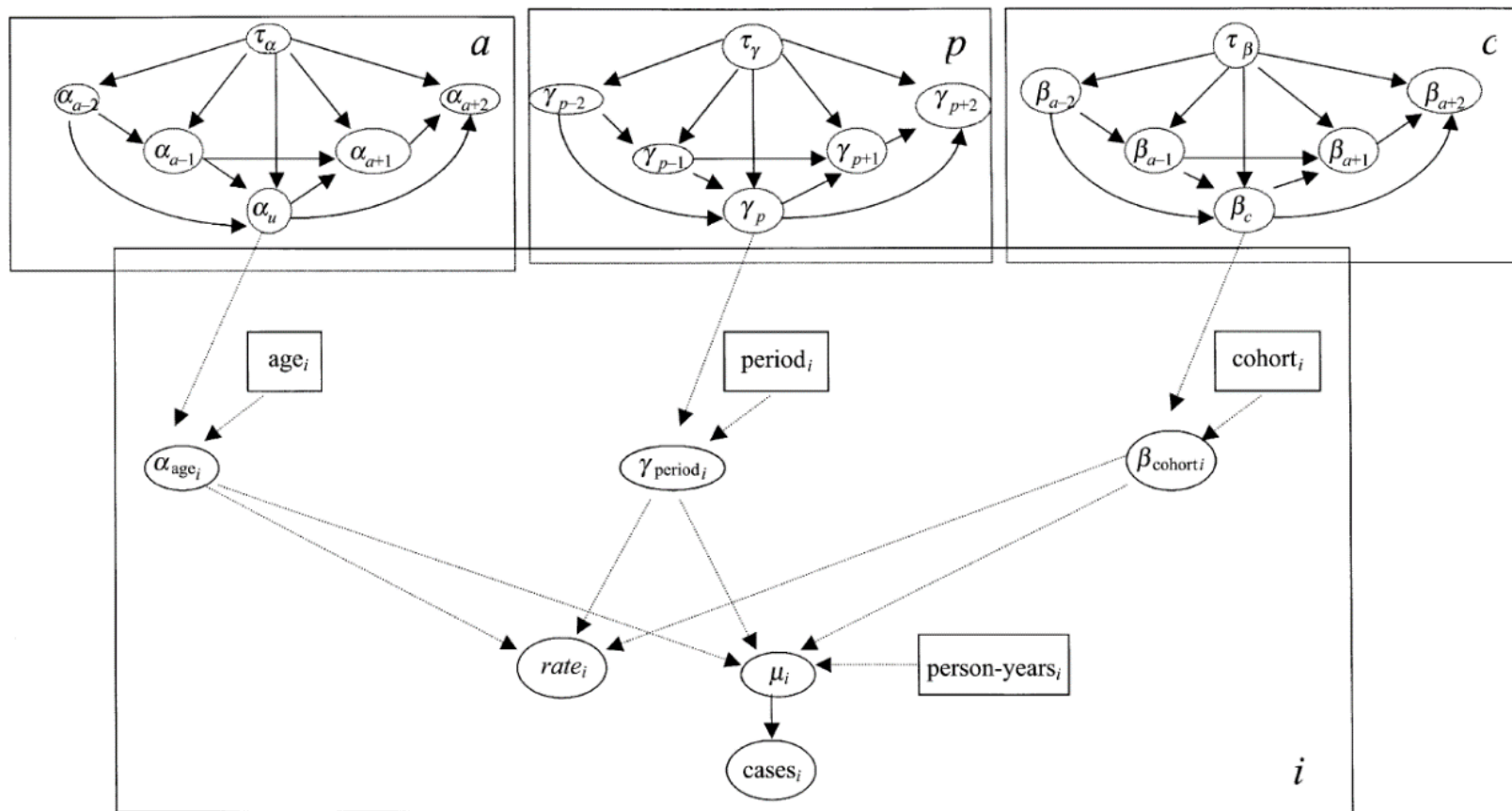
$$\lambda_i = \exp(\alpha_{età_i} + \beta_{periodo_i} + \gamma_{coorte_i}) \quad \text{per } 1 \leq i \leq A(P + N)$$

dove  $\lambda_i$  è il tasso per  $i$ .

I risultati dell’applicazione di questo modello, presentati in questa tesi nel capitolo seguente, sono basati su 25000 iterazioni, dopo un burn-in di 5000 iterazioni. Il numero di iterazioni da eseguire e di burn-in è stato determinato tramite il test di Raftery e Lewis [41]. La convergenza dei parametri e dei tassi stimati e predetti è stata valutata attraverso sia test grafici, come il trace-plot e il grafico dell’autocorrelazione, sia con test più formali come il test di Heidelberg e Welch e quello di Gelman e Rubin.

In Figura 5 viene mostrato un grafico aciclico diretto (tipico del software Winbugs) per questo modello per rappresentare le assunzioni di indipendenza condizionale. I nodi sono sia variabili (ovali) che costanti (rettangoli) e le dipendenze tra questi nodi sono logiche (freccie tratteggiate) o probabilistiche (freccie continue).

In questa tesi, le analisi del modello bayesiano sono state svolte con i software statistici Winbugs per quanto riguarda l’applicazione della tecnica MCMC e R per la parte rimanente.



**Figura 5:** Rappresentazione grafica del modello età-periodo-coorte bayesiano in cui i casi<sub>*i*</sub> hanno una distribuzione di Poisson con media  $\mu_i$ .  $\mu_i$  e i tassi<sub>*i*</sub> dipendono logicamente dalle altre variabili [37]

#### 4.4. Modello età-periodo bayesiano

Nel paragrafo precedente ho descritto il metodo di previsione dei dati di mortalità proposto da Bray [37], che si basa su un modello di analisi età-periodo-coorte con approccio bayesiano. Tale metodo può essere modificato per regredire un modello età-periodo (AP) per un numero di morti annuali e ottenere proiezioni a breve termine. A tal proposito è stato creato un modello AP basato su classi di età e periodi annuali [42]. Essendo i dati di mortalità utilizzati per l'applicazione in questa tesi basati su classi di età quinquennali, ho provveduto ad impostare un differente modello AP.

Questo metodo, in modo simile a quello di Bray, si basa su un modello log-lineare e assume un modello autoregressivo di secondo ordine, come distribuzioni a priori, per predire le funzioni di distribuzione di probabilità a posteriori dei tassi, facendo inferenza bayesiana con le tecniche computazionali MCMC.

Il modello è definito come segue.

Assumendo:

$$d_{ap} \sim \text{Poisson}(\mu_{ap})$$

con  $a=1, \dots, A$  e  $p=1, \dots, P$

Il modello età-periodo è definito come:

$$\log \mu_{ap} = \log m_{ap} + \alpha_a + \beta_p$$

dove  $\mu_{ap}$  è la media dei morti  $d_{ap}$ ,  $m_{ap}$  sono gli anni-persona a rischio per l' $a$ -esima classe di età al  $p$ -esimo anno di calendario,  $\alpha_a$  è l'effetto di età e  $\beta_p$  è l'effetto di periodo. Si supponga che siano richieste le proiezioni per i futuri  $N$  anni di calendario.

In modo analogo al modello APC descritto in precedenza, le distribuzioni a priori per i parametri del modello possono essere definite come segue.

Per gli  $A$  effetti di età:

$$\alpha_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\alpha}\right)$$



$$\alpha_2|\alpha_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\alpha}\right)$$

$$\alpha_a|\alpha_{1,\dots,a-1} \sim \text{Normale}\left(2\alpha_{a-1} - \alpha_{a-2}, \frac{1}{\tau_\alpha}\right) \quad \text{per } 3 \leq a \leq A$$

Per i P+N effetti di periodo:

$$\beta_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\beta}\right)$$

$$\beta_2|\beta_1 \sim \text{Normale}\left(0, 1000000 \frac{1}{\tau_\beta}\right)$$

$$\beta_p|\beta_{1,\dots,p-1} \sim \text{Normale}\left(2\beta_{p-1} - \beta_{p-2}, \frac{1}{\tau_\beta}\right) \quad \text{per } 3 \leq p \leq P + N$$

Agli iperparametri di precisione  $\tau_\alpha$  e  $\tau_\beta$  sono state assegnate delle distribuzioni iperpriori vaghe, ossia:

$$\tau_\alpha \sim \text{Gamma}(0.001, 0.001)$$

$$\tau_\beta \sim \text{Gamma}(0.001, 0.001)$$

Infine ricombinando gli effetti “smussati” di età e periodo si ottengono le stime e le proiezioni dei tassi nel seguente modo:

$$\lambda_{ap} = \exp(\alpha_a + \beta_p) \quad \text{per } 1 \leq p \leq P + N$$

dove  $\lambda_{ap}$  è il tasso.

In questa tesi si è applicato questo metodo a periodi annuali e quinquennali. Nel primo caso si sono eseguite 14000 iterazioni, dopo un burn-in di 1000 iterazioni, mentre nel caso dei periodi quinquennali 27000 iterazioni dopo un burn-in di 1000. Anche in questa analisi si è usato il test di Raftery and Lewis per determinare il numero di iterazioni da eseguire e di burn-in. La convergenza dei parametri e dei tassi stimati e predetti è stata valutata sia attraverso test grafici, come il trace-plot e il

grafico dell'autocorrelazione, sia con test più formali come il test di Heidelberg and Welch e quello di Gelman and Rubin.

Le analisi del modello bayesiano sono state svolte con Winbugs e R.

## 4.5. Test di confronto

Per confrontare le performance di questi metodi di previsione si calcola la deviazione percentuale assoluta (*absolute percentage deviation*, APD) tra le previsioni e i valori reali per ciascuna unità temporale di previsione. Questa misura è data dal valore assoluto della differenza tra il valore previsto e quello osservato diviso il valore osservato per cento [33]. In formula:

$$APD_i = \frac{|valore\ previsto_i - valore\ osservato_i|}{valore\ osservato_i} \times 100$$

per  $i=1, \dots, n$

dove  $i$  indica l'unità temporale di previsione.

Inoltre, si calcola la deviazione percentuale assoluta media (*average absolute percentage deviation*, AAPD), cioè la media di questi rapporti di errori di previsione su tutte le unità temporali di previsione.

$$AAPD = \frac{1}{n} \sum_{i=1}^n \frac{|valore\ previsto_i - valore\ osservato_i|}{valore\ osservato_i} \times 100$$

Valori piccoli dell'AAPD indicano che le stime previste sono vicine al valore osservato. Poiché una previsione viene considerata affidabile quando il valore dell'AAPD risulta essere inferiore al 5% [33], si provvede a comparare i modelli in termini di performance attraverso il conteggio di  $AAPD \leq 5\%$ .

Infine, si calcola la mediana degli AAPD per ciascun metodo di previsione.

# CAPITOLO 5

## Applicazione ai dati reali

---

In questo capitolo viene mostrata un'applicazione di alcuni dei metodi presentati nei capitoli precedenti su dati di mortalità per tumore.

### 5.1. Descrizione dei dati

I dati delle morti certificate per tumore, quando disponibili, sono stati ottenuti dal database di mortalità dell'OMS, consultabile su supporto elettronico (WHOSIS), per il periodo 1970-2009 [43]. I dati considerati sono relativi all'Unione Europea (UE) nel suo insieme, la quale è costituita da 27 Stati membri, come è stata definita al 1° gennaio 2007. Per i dettagli della costruzione del dataset dell'UE vedere Appendice 1.

I dati di mortalità sono codificati in base all'ICD (*International Classification of Diseases*), ossia la Classificazione Statistica Internazionale delle Malattie e dei Problemi sanitari correlati, proposta dall'OMS. Durante il periodo considerato in questo studio, sono state utilizzate tre diverse revisioni di questa classificazione (ICD-8, ICD-9 E ICD-10) [44-46]. Poiché le differenze nelle classificazioni tra le diverse revisioni erano minime, i decessi per tumore sono stati ricodificati secondo la decima revisione dell'ICD (ICD-10) [47, 48].

Le stime delle popolazioni residenti per i corrispettivi periodi, basate sui censimenti ufficiali, sono state estratte dallo stesso database dell'OMS [43].

Dalle matrici dei dati delle morti certificate e delle popolazioni residenti abbiamo calcolato il numero di decessi e i tassi di mortalità specifici per età per 100.000 abitanti per classi di età di 5 anni (da 0-4 a 85+ anni), per ciascun periodo (dal 1970 al 2009). Per la costruzione del dataset europeo, quando i dati sono mancanti, si sostituisce o si interpola con i dati disponibili più vicini.

L'applicazione dei metodi di analisi età-periodo-coorte considerati è stata eseguita sui dati della mortalità per tutti i tumori (ICD-10: C00-D48). Per questa analisi APC sono stati usati i tassi età-specifici di mortalità per 100.000 uomini per le classi di età quinquennali (da 20-24 a 75-79 anni) e per i periodi quinquennali considerati (da 1970-74 a 2005-09). Le coorti sono state definite in base al loro anno di nascita centrale. Quindi, la prima coorte possibile è quella del 1895 e si riferisce agli individui di età compresa tra i 75 e i 79 anni che sono morti nel quinquennio 1970-74, ma che potrebbero essere nati in uno qualsiasi dei dieci anni tra il 1890 e il 1899.

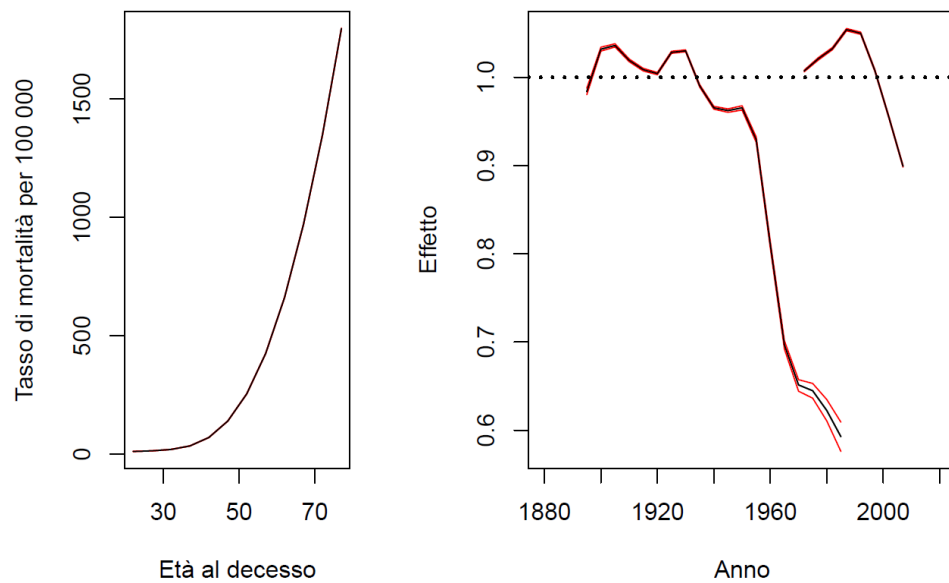
Invece, per quanto riguarda l'applicazione dei metodi di analisi predittiva considerati, si sono considerati i dati delle morti per i principali siti tumorali, ovvero stomaco (ICD-10: C16), intestino (ICD-10: C17-C21, C26), pancreas (ICD-10: C25), polmone (ICD-10: C33-C34), mammella (ICD-10: C50), utero (ICD-10: C53-C55), prostata (ICD-10: C61) e leucemia (ICD-10: C91-C95), e per tutti i tumori (ICD-10: C00-D48). Sono stati utilizzati il numero di decessi e i tassi età-specifici di mortalità per 100.000 uomini e donne per le classi di età quinquennali (da 0-4 a 85+ anni), per i periodi annuali (dal 1970 al 2009) e quinquennali (da 1970-74 a 2005-09).

## **5.2. Analisi età-periodo-coorte**

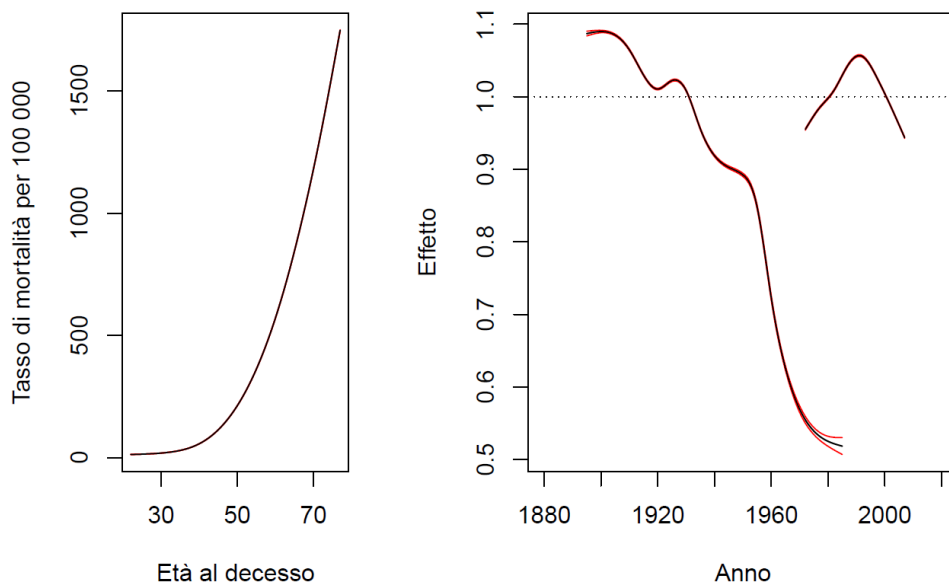
I dati descritti nel paragrafo precedente sono stati analizzati con tutti i tre metodi di analisi età-periodo-coorte considerati in questa tesi: metodo della verosimiglianza penalizzata, metodo dei GAM e metodo dei PLS.

### **5.2.1. Risultati**

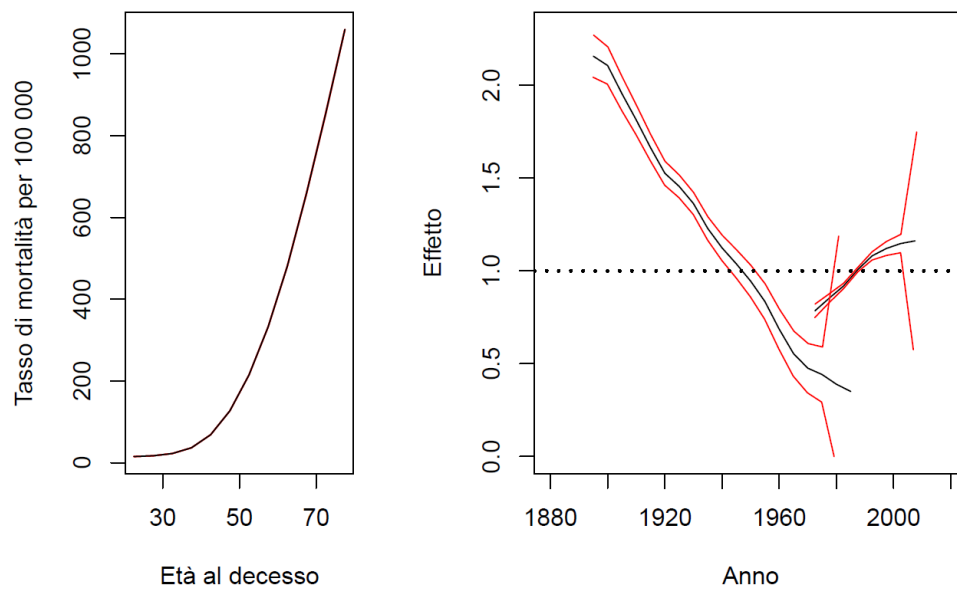
Dall'applicazione dei metodi di analisi età-periodo-coorte descritti nel terzo capitolo di questa tesi (metodo della verosimiglianza penalizzata, GAM e PLS) ai dati di mortalità per tutti i tumori negli uomini europei, si ottengono rispettivamente le seguenti rappresentazioni grafiche dei modelli:



**Figura 6:** Modello età-periodo-coorte calcolato usando il metodo della verosimiglianza penalizzata con intervalli di confidenza calcolati con la procedura bootstrap.



**Figura 7:** Modello età-periodo-coorte calcolato usando il metodo dei modelli additivi generalizzati con intervalli di confidenza calcolati con il metodo della verosimiglianza profilo.



**Figura 8:** Modello età-periodo-coorte calcolato usando il metodo dei minimi quadrati parziali con intervalli di confidenza calcolati con il metodo jackknife.

I grafici mostrati per i rispettivi metodi di analisi APC sono suddivisi in due riquadri. Il primo riquadro mostra la curva dell'effetto età (curva nera) con l'IC al 95% (curve rosse) derivante dall'analisi APC rispetto all'età, mentre il secondo riquadro illustra gli effetti della coorte di nascita e del periodo di morte (curve nere) con i corrispondenti IC al 95% (curve rosse).

Osservando le rappresentazioni grafiche delle analisi APC ottenute con il metodo della verosimiglianza penalizzata e con il metodo dei modelli additivi generalizzati si possono fare le seguenti considerazioni.

Le curve degli effetti età mettono in rilievo la forte dipendenza dall'età, con tassi per i gruppi di età più vecchi che raggiungono valori di 1800/100.000 uomini, come ci si aspetterebbe dalla patologia in studio. Il grafico dell'effetto di coorte per l'UE ha tre distinti valori massimi, che corrispondono ai picchi degli andamenti degli effetti di coorte dei diversi paesi principali, e un forte effetto decrescente nelle coorti più giovani dopo il 1960. L'UE presenta una struttura dell'effetto del periodo crescente fino alla fine degli anni '80/inizio degli anni '90, dove si ha un picco, e poi decresce notevolmente.

Gli andamenti della mortalità per tutti i tumori nei maschi dell'UE sono il risultato dei modelli applicati a tutta la popolazione europea per i diversi tumori. Essi riflettono in modo preponderante gli andamenti della mortalità per tumori legati al fumo, in particolare il cancro del polmone (principale causa di morte per tumore nei maschi europei), e i diversi stadi dell'epidemia del fumo nelle popolazioni costituenti l'Unione Europea, mostrando che il fumo è la principale causa di morte per tumore negli uomini europei [49]. Questo può essere chiaramente visto nei tre picchi presenti nella curva degli effetti di coorte dell'UE. Infatti, analizzando le diverse strutture delle curve degli effetti di coorte dei vari paesi europei (di cui non riporto i grafici in questa tesi), si nota come essi riflettono la diversa esposizione e controllo dei fattori di rischio (in particolare il fumo di tabacco) nelle successive generazioni di uomini nei vari paesi [50, 51]. Per quanto riguarda gli effetti di periodo, la curva favorevole negli ultimi decenni probabilmente riflette i cambiamenti e i modesti progressi della diagnosi e del trattamento [52].

Osservando i grafici dell'analisi APC ottenuti con il metodo dei minimi quadrati parziali, notiamo subito come questi siano notevolmente diversi rispetto a quelli degli altri due metodi.

Le curve degli effetti dell'età mostrano ancora che la mortalità dipende molto dall'età, ma la grandezza dei tassi per i gruppi di età più vecchi raggiunge valori molto inferiori (1060/100.000 uomini). Invece il grafico dell'effetto di coorte per l'UE è molto diverso, infatti ha un valore massimo per le coorti nate intorno al 1900 e un forte effetto decrescente nelle coorti nate successivamente. Inoltre i valori dei rischi relativi dell'effetto di coorte sono decisamente più forti. In ultimo, il grafico dell'effetto del periodo presenta una struttura crescente per tutto il periodo.

### **5.2.2. Confronto dei metodi statistici utilizzati**

A questo punto, è doveroso fare alcune considerazioni generali sul modello APC. Innanzitutto, le stime degli effetti dell'età, periodo e coorte presentano diversi problemi di variabilità casuale. Per le stime dell'età il problema tende a manifestarsi in particolare per i gruppi di età più giovani, dove il numero assoluto di morti è basso

per la maggior parte delle neoplasie. Negli effetti di coorte questi problemi sono più grandi per entrambe le estremità della curva: come viene evidenziato dagli intervalli di confidenza, le prime e le ultime coorti di nascita si basano su un numero limitato di osservazioni, mentre andando verso le coorti centrali il numero di osservazioni aumenta. Inoltre, le coorti più recenti hanno un numero molto piccolo di morti, poiché rappresentano le classi di età più giovani. Ne consegue che i cambiamenti degli andamenti in queste coorti recenti dovrebbero essere interpretati con cautela, come è mostrato dagli ampi intervalli di confidenza per le loro stime, ciò nonostante forniscono importanti informazioni per gli andamenti futuri.

Dai risultati dei modelli APC eseguiti, si nota che i primi due metodi (verosimiglianza penalizzata e GAM con splines) producono risultati simili, mentre il terzo metodo (PLS) presenta una stima dei parametri profondamente diversa. Possiamo cercare la spiegazione di queste differenze nelle modalità che i tre modelli utilizzano per ovviare al problema della collinearità perfetta dei tre parametri età, periodo e coorte. Da una parte il metodo della verosimiglianza penalizzata e quello con i modelli GAM utilizzano tecniche differenti per distribuire l'effetto della deriva temporale tra le coorti e i periodi nel modo più corretto possibile. Dall'altra parte il modello stimato con il metodo dei PLS risolve il problema dell'identificabilità dal punto di vista della risoluzione dell'inversa generalizzata, in modo da minimizzare la matrice di varianze e covarianze tra i possibili parametri stimati.

### **5.3. Analisi predittiva**

Per quanto concerne l'analisi predittiva, si sono applicati ai dati reali e confrontati solo alcuni dei metodi presentati nel quarto capitolo di questa tesi. I metodi che si sono usati per le proiezioni dei dati di nostro interesse sono: modello età-periodo con joinpoint sul numero di morti (con modello di regressione lineare, di Poisson e log-lineare), modello età-periodo bayesiano e modello età-periodo-coorte bayesiano.



### **5.3.1. Risultati**

I metodi di proiezione della mortalità sono stati applicati ai dati di mortalità per i principali siti tumorali e per tutti i tumori per gli individui residenti nell'UE, nel periodo compreso tra il 1970 e il 1999, con l'obiettivo di predire i dati per il periodo 2000-2009. Oltre alle previsioni per singolo anno di calendario, si sono effettuate anche predizioni della mortalità per periodi quinquennali, al fine di poter implementare il modello età-periodo-coorte bayesiano. Infatti, i modelli età-periodo-coorte vengono solitamente implementati con classi di età e periodi di 5 anni.

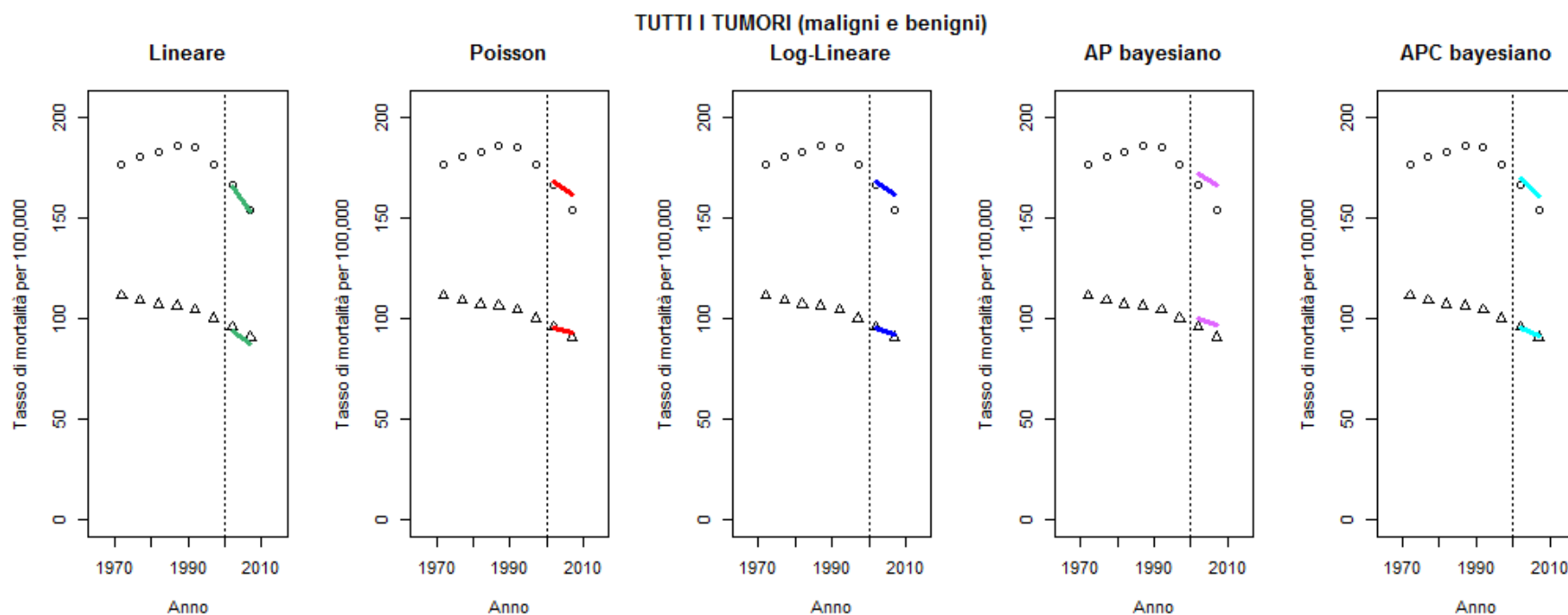
Di seguito, mostrerò i risultati ottenuti dalle applicazioni rispettivamente per periodi quinquennali e annuali.

#### **5.3.1.1. Periodi quinquennali**

Di seguito mostriamo la tabella per i tassi di mortalità osservati e predetti con il metodo età-periodo con joinpoint sul numero di morti e attraverso i metodi età-periodo e età-periodo-coorte bayesiani, per tutti i tumori per il periodo 2000-2009, suddiviso in quinquenni (2000-04, 2005-09), nell'Unione Europea (Tabella 5). Queste previsioni vengono anche rappresentate graficamente in Figura 9.

| TUTTI I TUMORI (maligni e benigni) |        |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
|------------------------------------|--------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|---------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso                              |        |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
| Maschi                             |        |                             |                             |                                     |                                     |                                      | Femmine |                             |                             |                                     |                                     |                                      |
|                                    | Oss.   | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |         | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| 2000-04                            | 166.15 | 165.41                      | 168.05                      | 168.00                              | 171.92                              | 169.04                               | 95.63   | 93.91                       | 95.43                       | 95.37                               | 99.66                               | 95.71                                |
| 2005-09                            | 153.50 | 152.64                      | 161.41                      | 161.30                              | 165.87                              | 160.48                               | 90.80   | 87.40                       | 92.72                       | 92.61                               | 96.96                               | 91.12                                |

**Tabella 5:** Tassi di mortalità osservati e predetti secondo cinque diversi modelli di previsione per periodi quinquennali.



**Figura 9:** Tassi di mortalità osservati e predetti secondo cinque diversi modelli di previsione per maschi (cerchi) e femmine (triangoli) per periodi quinquennali.

Osservando la tabella e i grafici si può notare che le previsioni con i cinque diversi modelli sono abbastanza simili, specialmente per il modello di Poisson e quello log-lineare. Tuttavia, il modello lineare tende a sottostimare i dati predetti, mentre gli altri quattro tendono a sovrastimare, soprattutto per il quinquennio di previsione più lontano nel tempo.

Delle osservazioni simili possono essere fatte anche sui risultati dei principali siti tumorali (stomaco, intestino, pancreas, polmone, mammella, utero, prostata e leucemia) analizzati, che sono riportati in Appendice 2.

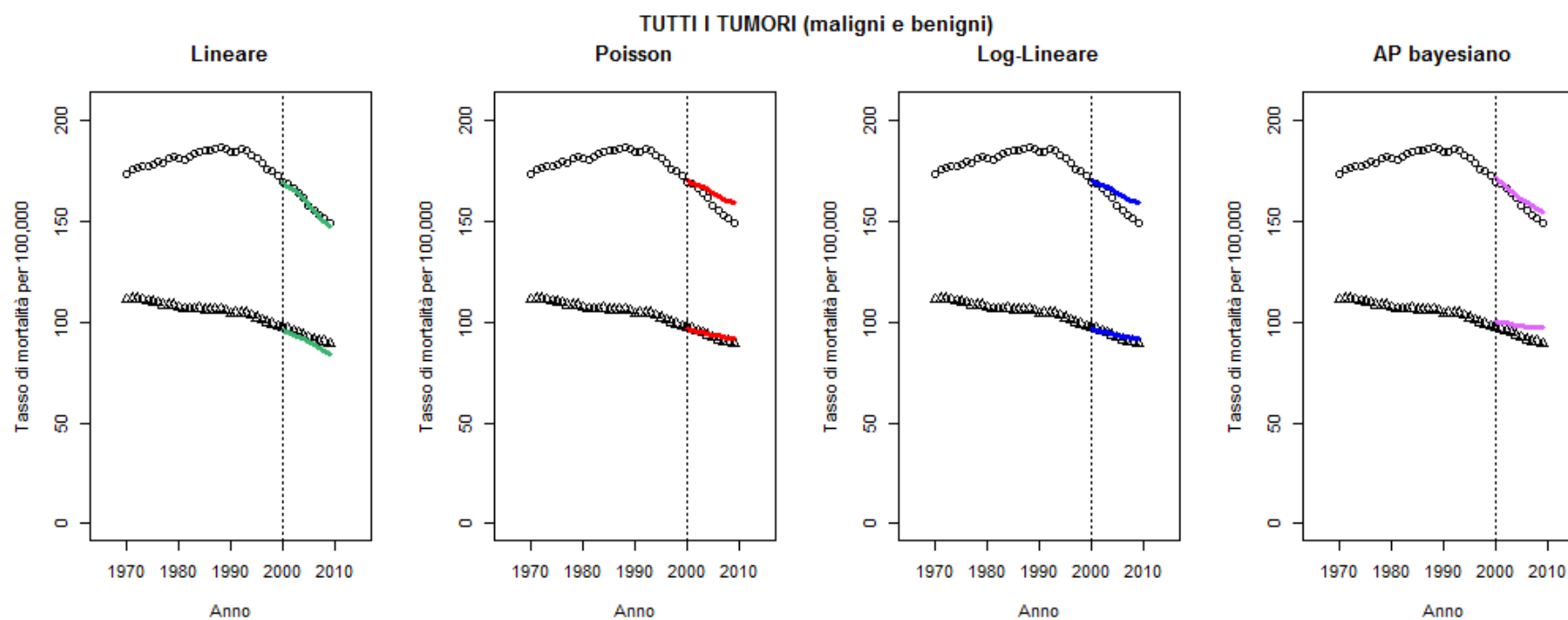
#### **5.3.1.2. Periodi annuali**

La Tabella 6 mostra i tassi di mortalità osservati e predetti con il metodo età-periodo con joinpoint sul numero di morti e il metodo età-periodo bayesiano, per tutti i tumori per i singoli anni del periodo 2000-2009 nell'Unione Europea. Queste previsioni sono rappresentate graficamente in Figura 10.

Anche in questo caso, osservando la tabella e i grafici, si può notare che le previsioni con i quattro diversi modelli sono abbastanza simili, specialmente per il modello di Poisson e quello log-lineare. Il modello lineare tende a sottostimare i dati predetti, mentre gli altri tre tendono a sovrastimare, soprattutto per gli anni di previsione più lontani nel tempo. Tutti i modelli, in accordo con i dati osservati, mostrano come la mortalità per tumore sia costantemente in diminuzione negli anni in entrambi i sessi. Delle osservazioni simili possono essere fatte anche sui risultati dei principali siti tumorali (stomaco, intestino, pancreas, polmone, mammella, utero, prostata e leucemia) analizzati, che sono riportati in Appendice 2.

| <b>TUTTI I TUMORI (maligni e benigni)</b> |             |                                      |                                      |  |   |                |                                      |                                      |  |   |
|---|-------------|--------------------------------------|--------------------------------------|--|---|----------------|--------------------------------------|--------------------------------------|--|---|
| <b>Sesso</b>                              |             |                                      |                                      |  |   |                |                                      |                                      |  |   |
| <b>Maschi</b>                             |             |                                      |                                      |  |   | <b>Femmine</b> |                                      |                                      |  |   |
|   | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> | <b>Oss.</b>    | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |
| <b>2000</b>                               | 169.61      | 169.77                               | 170.70                               | 170.67                                   | 172.35                                    | 97.44          | 96.31                                | 96.86                                | 96.82                                    | 100.66                                    |
| <b>2001</b>                               | 168.48      | 167.43                               | 169.05                               | 169.01                                   | 169.88                                    | 96.58          | 95.06                                | 95.99                                | 95.95                                    | 100.05                                    |
| <b>2002</b>                               | 166.38      | 165.48                               | 167.93                               | 167.88                                   | 167.52                                    | 95.59          | 93.92                                | 95.33                                | 95.28                                    | 99.48                                     |
| <b>2003</b>                               | 164.51      | 163.85                               | 167.28                               | 167.22                                   | 165.27                                    | 94.95          | 93.01                                | 94.99                                | 94.93                                    | 98.96                                     |
| <b>2004</b>                               | 162.03      | 161.97                               | 166.52                               | 166.45                                   | 163.14                                    | 93.65          | 92.05                                | 94.70                                | 94.62                                    | 98.49                                     |
| <b>2005</b>                               | 157.95      | 158.39                               | 164.11                               | 164.03                                   | 161.17                                    | 92.54          | 90.46                                | 93.84                                | 93.75                                    | 98.09                                     |
| <b>2006</b>                               | 155.75      | 155.74                               | 162.83                               | 162.73                                   | 159.36                                    | 91.26          | 89.02                                | 93.25                                | 93.15                                    | 97.76                                     |
| <b>2007</b>                               | 153.46      | 152.64                               | 161.19                               | 161.08                                   | 157.72                                    | 90.66          | 87.40                                | 92.59                                | 92.48                                    | 97.53                                     |
| <b>2008</b>                               | 151.62      | 150.04                               | 160.25                               | 160.12                                   | 156.27                                    | 90.42          | 85.96                                | 92.23                                | 92.10                                    | 97.39                                     |
| <b>2009</b>                               | 149.05      | 147.30                               | 159.28                               | 159.13                                   | 155.03                                    | 89.19          | 84.48                                | 91.94                                | 91.80                                    | 97.34                                     |

**Tabella 6:** Tassi di mortalità osservati e predetti secondo quattro diversi modelli di previsione per periodi annuali.



**Figura 10:** Tassi di mortalità osservati e predetti secondo quattro diversi modelli di previsione per maschi (cerchi) e femmine (triangoli) per periodi annuali.

### **5.3.2. Confronto dei metodi statistici utilizzati**

Di seguito, confronterò e discuterò i metodi di predizione considerati nell'applicazione di questa tesi rispettivamente per i periodi quinquennali e annuali.

#### **5.3.2.1. Periodi quinquennali**

In Tabella 7 sono riportati i valori dell'AAPD per i cinque modelli predittivi considerati per i siti tumorali analizzati, la loro mediana totale e il numero di volte in cui il valore dell'AAPD risulta inferiore al 5%.

Osservando i valori degli AAPD, notiamo che le previsioni prodotte dai modelli analizzati sono tutte abbastanza affidabili, poiché si hanno valori percentuali inferiori al 5% per la maggior parte dei risultati. Ci sono alcune eccezioni. Il modello lineare ha degli AAPD elevati per le previsioni dei dati di mortalità per tumore dello stomaco, ma ciò potrebbe essere una conseguenza del fatto che questa neoplasia presenta un andamento della mortalità esponenzialmente asintotico che si adatta a curve di tipo logaritmico o esponenziali come Poisson e log-lineari. Lo stesso vale per il tumore del polmone nelle femmine. Invece, per le leucemie, che hanno una mortalità bassa, sembrerebbe che il modello lineare e il modello bayesiano basato su un'analisi età-periodo-coorte abbiano delle stime più affidabili. Le previsioni effettuate con i modelli di Poisson, log-lineare e età-periodo bayesiano per il tumore dell'intestino, sebbene simili, sembrerebbero meno attendibili rispetto a quelle degli altri due modelli. In particolare per le femmine, dal 1980 si osserva un decremento dei tassi compatibile con un andamento lineare inverso; ciò giustifica l'AAPD del modello lineare più basso rispetto a quello degli altri modelli. I due modelli con approccio bayesiano hanno valori dell'AAPD molto elevati per il tumore della prostata. Il modello AP bayesiano non predirebbe bene i tassi di mortalità del tumore della mammella femminile.

|                           | Sesso                      |                            |                                |                                 |                                  |                            |                            |                                |                                 |                                  |
|---------------------------|----------------------------|----------------------------|--------------------------------|---------------------------------|----------------------------------|----------------------------|----------------------------|--------------------------------|---------------------------------|----------------------------------|
|                           | Maschi                     |                            |                                |                                 |                                  | Femmine                    |                            |                                |                                 |                                  |
|                           | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-Lineare | AAPD<br>Modello AP<br>Bayesiano | AAPD<br>Modello APC<br>Bayesiano | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-Lineare | AAPD<br>Modello AP<br>Bayesiano | AAPD<br>Modello APC<br>Bayesiano |
| <b>STOMACO</b>            | 6.05                       | 1.19                       | 1.08                           | 0.40                            | 1.00                             | 8.23                       | 1.53                       | 1.49                           | 1.93                            | 3.42                             |
| <b>INTESTINO</b>          | 2.50                       | 6.60                       | 6.46                           | 6.58                            | 4.44                             | 0.24                       | 5.24                       | 5.22                           | 6.26                            | 3.17                             |
| <b>PANCREAS</b>           | 2.84                       | 0.52                       | 0.58                           | 3.17                            | 4.40                             | 2.46                       | 2.18                       | 2.26                           | 1.90                            | 4.61                             |
| <b>POLMONE</b>            | 3.91                       | 0.43                       | 0.42                           | 5.85                            | 3.97                             | 9.67                       | 4.94                       | 5.03                           | 1.48                            | 3.79                             |
| <b>MAMMELLA</b>           | .                          | .                          | .                              | .                               | .                                | 5.58                       | 2.06                       | 2.04                           | 9.73                            | 3.34                             |
| <b>UTERO</b>              | .                          | .                          | .                              | .                               | .                                | 2.29                       | 3.28                       | 3.19                           | 2.85                            | 0.62                             |
| <b>PROSTATA</b>           | 1.32                       | 5.86                       | 5.74                           | 13.56                           | 13.37                            | .                          | .                          | .                              | .                               | .                                |
| <b>LEUCEMIA</b>           | 1.84                       | 5.11                       | 5.13                           | 9.57                            | 2.06                             | 2.92                       | 7.51                       | 7.62                           | 21.88                           | 1.30                             |
| <b>TUTTI I<br/>TUMORI</b> | 0.50                       | 3.15                       | 3.10                           | 5.76                            | 3.14                             | 2.77                       | 1.16                       | 1.13                           | 5.50                            | 0.22                             |
| <b>MEDIANA<br/>TOTALE</b> | 2.50                       | 3.15                       | 3.10                           | 5.85                            | 3.97                             | 2.84                       | 2.73                       | 2.73                           | 4.17                            | 3.25                             |
| <b>AAPD &lt; 5%</b>       | 6                          | 4                          | 4                              | 2                               | 6                                | 5                          | 6                          | 5                              | 4                               | 8                                |

**Tabella 7:** Media della deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare, AP e APC bayesiani per ciascun tumore per maschi e femmine per il periodo 2000-09. Mediana degli AAPD per ciascun modello di previsione. Numero di AAPD <5% per ciascun modello.



Tuttavia, osservando le mediane degli AAPD sembrerebbe che questi modelli predittivi forniscano delle predizioni abbastanza affidabili eccetto per il modello AP bayesiano nei maschi. Il conteggio degli AAPD minori del 5% mostra che complessivamente il modello che sembra predire meglio in entrambi i sessi per il periodo 2000-2009 sia il modello APC bayesiano.

Detto ciò, confrontiamo i modelli di previsione nel breve e medio periodo.

Nella Tabella 8, che presenta i valori degli APD per il periodo 2000-04, notiamo che restringendo le analisi al breve periodo, le proiezioni della mortalità sembrano migliorare molto. Tuttavia, i valori degli APD restano elevati per il tumore della prostata in entrambi i modelli bayesiani e per la leucemia nel modello AP bayesiano. Nel complesso le mediane degli APD risultano inferiori al 5% per tutti i modelli, quindi questi modelli risultano affidabili per proiezioni della mortalità nel breve periodo.

La Tabella 9 riporta i valori degli APD di ciascun modello per il periodo 2005-09. Osserviamo che i modelli analizzati predirebbero meno bene nel medio periodo. In particolare, per il tumore della prostata e quello del colon-retto, solo il modello lineare ha valori degli APD minori del 5%, mentre per la leucemia il modello lineare e quello APC bayesiano presentano una migliore performance di predizione. Al contrario, il modello lineare non predirebbe molto bene la mortalità del tumore dello stomaco. Il tumore del polmone presenta valori degli APD abbastanza contrastanti per i due sessi. Complessivamente, dalle mediane possiamo dedurre che il modello che risulta meno adatto a predire dati di mortalità nel medio periodo sia il modello AP bayesiano.

|                           | Sesso                     |                           |                               |                                |                                 |                           |                           |                               |                                |                                 |
|---------------------------|---------------------------|---------------------------|-------------------------------|--------------------------------|---------------------------------|---------------------------|---------------------------|-------------------------------|--------------------------------|---------------------------------|
|                           | Maschi                    |                           |                               |                                |                                 | Femmine                   |                           |                               |                                |                                 |
|                           | APD<br>Modello<br>Lineare | APD<br>Modello<br>Poisson | APD<br>Modello<br>Log-Lineare | APD<br>Modello AP<br>Bayesiano | APD<br>Modello APC<br>Bayesiano | APD<br>Modello<br>Lineare | APD<br>Modello<br>Poisson | APD<br>Modello<br>Log-Lineare | APD<br>Modello AP<br>Bayesiano | APD<br>Modello APC<br>Bayesiano |
| <b>STOMACO</b>            | 3.25                      | 0.33                      | 0.23                          | 0.32                           | 0.72                            | 4.88                      | 0.61                      | 0.80                          | 1.68                           | 2.03                            |
| <b>INTESTINO</b>          | 1.58                      | 3.32                      | 3.23                          | 3.57                           | 2.48                            | 0.38                      | 2.39                      | 2.37                          | 4.46                           | 2.14                            |
| <b>PANCREAS</b>           | 0.74                      | 0.79                      | 0.81                          | 1.44                           | 2.18                            | 0.68                      | 1.30                      | 1.33                          | 0.33                           | 2.56                            |
| <b>POLMONE</b>            | 2.08                      | 0.61                      | 0.60                          | 4.22                           | 3.08                            | 5.42                      | 3.20                      | 3.29                          | 1.08                           | 1.94                            |
| <b>MAMMELLA</b>           | .                         | .                         | .                             | .                              | .                               | 3.36                      | 1.92                      | 1.91                          | 7.17                           | 2.53                            |
| <b>UTERO</b>              | .                         | .                         | .                             | .                              | .                               | 0.49                      | 2.05                      | 1.97                          | 3.65                           | 0.32                            |
| <b>PROSTATA</b>           | 0.31                      | 2.22                      | 2.13                          | 6.87                           | 6.71                            | .                         | .                         | .                             | .                              | .                               |
| <b>LEUCEMIA</b>           | 2.11                      | 1.84                      | 1.83                          | 6.84                           | 2.45                            | 3.59                      | 2.67                      | 2.75                          | 16.92                          | 0.14                            |
| <b>TUTTI I<br/>TUMORI</b> | 0.45                      | 1.14                      | 1.11                          | 3.47                           | 1.74                            | 1.80                      | 0.21                      | 0.27                          | 4.22                           | 0.08                            |
| <b>MEDIANA<br/>TOTALE</b> | 1.58                      | 1.14                      | 1.11                          | 3.57                           | 2.45                            | 2.58                      | 1.99                      | 1.94                          | 3.93                           | 1.98                            |

**Tabella 8:** Deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare, AP e APC bayesiani per ciascun tumore per maschi e femmine per il quinquennio 2000-04. Mediana degli APD per ciascun modello di previsione.

|                           | Sesso                     |                           |                               |                                |                                 |                           |                           |                               |                                |                                 |
|---------------------------|---------------------------|---------------------------|-------------------------------|--------------------------------|---------------------------------|---------------------------|---------------------------|-------------------------------|--------------------------------|---------------------------------|
|                           | Maschi                    |                           |                               |                                |                                 | Femmine                   |                           |                               |                                |                                 |
|                           | APD<br>Modello<br>Lineare | APD<br>Modello<br>Poisson | APD<br>Modello<br>Log-Lineare | APD<br>Modello AP<br>Bayesiano | APD<br>Modello APC<br>Bayesiano | APD<br>Modello<br>Lineare | APD<br>Modello<br>Poisson | APD<br>Modello<br>Log-Lineare | APD<br>Modello AP<br>Bayesiano | APD<br>Modello APC<br>Bayesiano |
| <b>STOMACO</b>            | 8.86                      | 2.04                      | 1.92                          | 0.47                           | 1.29                            | 11.59                     | 2.45                      | 2.18                          | 2.18                           | 4.82                            |
| <b>INTESTINO</b>          | 3.42                      | 9.89                      | 9.69                          | 9.59                           | 6.40                            | 0.09                      | 8.08                      | 8.06                          | 8.06                           | 4.19                            |
| <b>PANCREAS</b>           | 4.94                      | 0.25                      | 0.35                          | 4.89                           | 6.63                            | 4.23                      | 3.06                      | 3.20                          | 3.47                           | 6.66                            |
| <b>POLMONE</b>            | 5.74                      | 0.25                      | 0.24                          | 7.48                           | 4.87                            | 13.91                     | 6.68                      | 6.78                          | 1.89                           | 5.64                            |
| <b>MAMMELLA</b>           | .                         | .                         | .                             | .                              | .                               | 7.79                      | 2.20                      | 2.16                          | 12.30                          | 4.15                            |
| <b>UTERO</b>              | .                         | .                         | .                             | .                              | .                               | 4.10                      | 4.50                      | 4.41                          | 2.05                           | 0.92                            |
| <b>PROSTATA</b>           | 2.34                      | 9.51                      | 9.35                          | 20.25                          | 20.03                           | .                         | .                         | .                             | .                              | .                               |
| <b>LEUCEMIA</b>           | 1.58                      | 8.39                      | 8.44                          | 12.29                          | 1.68                            | 2.24                      | 12.34                     | 12.49                         | 26.85                          | 2.47                            |
| <b>TUTTI I<br/>TUMORI</b> | 0.56                      | 5.16                      | 5.08                          | 8.06                           | 4.55                            | 3.75                      | 2.12                      | 1.99                          | 6.78                           | 0.35                            |
| <b>MEDIANA<br/>TOTALE</b> | 3.42                      | 5.16                      | 5.08                          | 8.06                           | 4.87                            | 4.17                      | 3.78                      | 3.81                          | 5.13                           | 4.17                            |

**Tabella 9:** Deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare, AP e APC bayesiani per ciascun tumore per maschi e femmine per il quinquennio 2005-09. Mediana degli APD per ciascun modello di previsione.

Infine, la seguente tabella (Tabella 10) mostra le mediane degli AAPD per tutti i modelli per il periodo 2000-2009.

|                | <i>Mediana AAPD</i>    |                        |                            |                             |                              |
|----------------|------------------------|------------------------|----------------------------|-----------------------------|------------------------------|
|                | <b>Modello Lineare</b> | <b>Modello Poisson</b> | <b>Modello Log-Lineare</b> | <b>Modello AP Bayesiano</b> | <b>Modello APC Bayesiano</b> |
| <b>TOTALE</b>  | 2.77                   | 3.15                   | 3.10                       | 5.76                        | 3.34                         |
| <b>Maschi</b>  | 2.50                   | 3.15                   | 3.10                       | 5.85                        | 3.97                         |
| <b>Femmine</b> | 2.84                   | 2.73                   | 2.73                       | 4.17                        | 3.25                         |

**Tabella 10:** Mediana degli AAPD per ciascun modello di previsione per il periodo 2000-09.

Osservando le mediane totali per i modelli di previsione si può notare che sono abbastanza simili, tranne che per il modello AP bayesiano (Mediana AAPD=5.76). Questo indica che il modello bayesiano basato su un'analisi età-periodo sembra risultare meno idoneo a predire la mortalità nel breve-medio periodo rispetto agli altri quattro modelli, che non mostrano grandi differenze di performance tra di loro.

Invece, se analizziamo le mediane separatamente per i due sessi, osserviamo che nei maschi il modello lineare risulta essere il migliore, mentre nelle femmine risultano migliori i tre modelli età-periodo con joinpoint (lineare, Poisson e log-lineare), i quali presentano una mediana di AAPD inferiore rispetto ai modelli bayesiani.

Da questi confronti, sembrerebbe che il modello lineare sia migliore nel predire i dati di mortalità rispetto agli altri quattro modelli.

### 5.3.2.2. Periodi annuali

Confrontiamo ora i modelli predittivi applicati a periodi annuali (anni di calendario). La Tabella 11 riporta i valori dell'AAPD per i quattro modelli di regressione considerati, vale a dire il modello lineare, quello di Poisson, quello log-lineare e il modello AP bayesiano, per i siti tumorali analizzati per il periodo 2000-09, la loro mediana totale e il numero di volte in cui il valore dell'AAPD risulta inferiore al 5%.

|                           | Sesso                      |                            |                                    |                                    |                            |                            |                                    |                                    |
|---------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|
|                           | Maschi                     |                            |                                    |                                    | Femmine                    |                            |                                    |                                    |
|                           | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano |
|                           |                            |                            |                                    |                                    |                            |                            |                                    |                                    |
| <b>STOMACO</b>            | 5.96                       | 1.41                       | 1.33                               | 4.30                               | 8.16                       | 1.47                       | 1.43                               | 5.36                               |
| <b>INTESTINO</b>          | 2.62                       | 6.71                       | 6.56                               | 3.98                               | 0.82                       | 5.38                       | 5.36                               | 8.40                               |
| <b>PANCREAS</b>           | 2.73                       | 0.70                       | 0.73                               | 3.15                               | 2.38                       | 2.30                       | 2.38                               | 9.97                               |
| <b>POLMONE</b>            | 3.84                       | 0.48                       | 0.47                               | 0.46                               | 9.52                       | 4.85                       | 4.94                               | 1.65                               |
| <b>MAMMELLA</b>           | .                          | .                          | .                                  | .                                  | 5.53                       | 2.00                       | 1.97                               | 9.52                               |
| <b>UTERO</b>              | .                          | .                          | .                                  | .                                  | 2.25                       | 3.37                       | 3.30                               | 4.70                               |
| <b>PROSTATA</b>           | 1.51                       | 6.03                       | 5.91                               | 5.49                               | .                          | .                          | .                                  | .                                  |
| <b>LEUCEMIA</b>           | 1.74                       | 5.32                       | 5.34                               | 11.42                              | 2.83                       | 7.61                       | 7.73                               | 22.03                              |
| <b>TUTTI I<br/>TUMORI</b> | 0.47                       | 3.24                       | 3.19                               | 1.85                               | 2.68                       | 1.35                       | 1.29                               | 5.79                               |
| <b>MEDIANA<br/>TOTALE</b> | 2.62                       | 3.24                       | 3.19                               | 3.98                               | 2.75                       | 2.84                       | 2.84                               | 7.10                               |
| <b>AAPD &lt; 5%</b>       | 6                          | 4                          | 4                                  | 5                                  | 5                          | 6                          | 6                                  | 2                                  |

**Tabella 11:** Media della deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare e AP bayesiano per ciascun tumore per maschi e femmine per il periodo 2000-09. Mediana degli AAPD per ciascun modello di previsione. Numero di AAPD <5% per ciascun modello.

I valori degli AAPD confermano quanto osservato precedentemente per i periodi quinquennali, cioè che le previsioni prodotte da questi modelli sono abbastanza affidabili, ma con alcune eccezioni. Il modello lineare ha degli AAPD elevati per i tassi di mortalità predetti per il tumore dello stomaco e per quello del polmone nelle femmine, mentre gli altri tre modelli presentano valori degli AAPD elevati per il tumore dell'intestino, quello della prostata e la leucemia. Dal conteggio degli AAPD minori del 5%, sembrerebbe che nei maschi il modello che predica meglio sia quello lineare, mentre nelle femmine i modelli di Poisson e log-lineare.

Gli stessi indicatori della Tabella 11 sono stati calcolati per il breve periodo (dal 2000 al 2004) e mostrati in Tabella 12.

|                           | Sesso                      |                            |                                    |                                    |                            |                            |                                    |                                    |
|---------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|
|                           | Maschi                     |                            |                                    |                                    | Femmine                    |                            |                                    |                                    |
|                           | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano |
|                           |                            |                            |                                    |                                    |                            |                            |                                    |                                    |
| <b>STOMACO</b>            | 3.08                       | 0.70                       | 0.67                               | 1.98                               | 4.72                       | 0.42                       | 0.61                               | 4.50                               |
| <b>INTESTINO</b>          | 1.75                       | 3.48                       | 3.38                               | 1.60                               | 0.63                       | 2.61                       | 2.58                               | 5.21                               |
| <b>PANCREAS</b>           | 0.62                       | 0.88                       | 0.90                               | 1.55                               | 0.62                       | 1.51                       | 1.53                               | 5.61                               |
| <b>POLMONE</b>            | 1.99                       | 0.52                       | 0.52                               | 0.53                               | 5.26                       | 3.08                       | 3.16                               | 0.60                               |
| <b>MAMMELLA</b>           | .                          | .                          | .                                  | .                                  | 3.27                       | 1.83                       | 1.82                               | 6.00                               |
| <b>UTERO</b>              | .                          | .                          | .                                  | .                                  | 0.44                       | 2.19                       | 2.13                               | 4.29                               |
| <b>PROSTATA</b>           | 0.61                       | 2.48                       | 2.40                               | 1.27                               | .                          | .                          | .                                  | .                                  |
| <b>LEUCEMIA</b>           | 1.96                       | 2.18                       | 2.18                               | 7.59                               | 3.47                       | 2.83                       | 2.91                               | 16.55                              |
| <b>TUTTI I<br/>TUMORI</b> | 0.34                       | 1.27                       | 1.24                               | 0.86                               | 1.65                       | 0.53                       | 0.54                               | 4.07                               |
| <b>MEDIANA<br/>TOTALE</b> | 1.75                       | 1.27                       | 1.24                               | 1.55                               | 2.46                       | 2.01                       | 1.97                               | 4.86                               |
| <b>AAPD &lt; 5%</b>       | 7                          | 7                          | 7                                  | 6                                  | 7                          | 8                          | 8                                  | 4                                  |

**Tabella 12:** Media della deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare e AP bayesiano per ciascun tumore per maschi e femmine per il periodo 2000-04. Mediana degli AAPD per ciascun modello di previsione. Numero di AAPD <5% per ciascun modello.

Considerando solo il breve periodo, secondo il confronto degli AAPD, questi modelli mostrano delle buone performance di proiezione della mortalità, ad eccezione del modello AP bayesiano per la leucemia. Complessivamente, le mediane degli AAPD risultano inferiori al 5% per tutti i modelli, pertanto questi modelli sono risultati essere affidabili per predire la mortalità nel breve periodo. Anche la frequenza di AAPD < 5% indica che i modelli sono quasi sempre appropriati, tranne che il modello AP bayesiano nelle femmine.

Passando ad analizzare il medio periodo (dal 2005 al 2009) (Tabella 13), si possono fare considerazioni simili a quanto osservato precedentemente per le analisi con periodi quinquennali. Infatti, questi modelli sembrerebbero meno appropriati a fare predizioni della mortalità nel medio periodo, soprattutto il modello AP bayesiano.

|                           | Sesso                      |                            |                                    |                                    |                            |                            |                                    |                                    |
|---------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|----------------------------|----------------------------|------------------------------------|------------------------------------|
|                           | Maschi                     |                            |                                    |                                    | Femmine                    |                            |                                    |                                    |
|                           | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano | AAPD<br>Modello<br>Lineare | AAPD<br>Modello<br>Poisson | AAPD<br>Modello<br>Log-<br>Lineare | AAPD<br>Modello<br>AP<br>Bayesiano |
|                           |                            |                            |                                    |                                    |                            |                            |                                    |                                    |
| <b>STOMACO</b>            | 8.85                       | 2.12                       | 1.99                               | 6.62                               | 11.61                      | 2.52                       | 2.24                               | 6.22                               |
| <b>INTESTINO</b>          | 3.49                       | 9.94                       | 9.75                               | 6.36                               | 1.01                       | 8.15                       | 8.14                               | 11.59                              |
| <b>PANCREAS</b>           | 4.84                       | 0.51                       | 0.56                               | 4.76                               | 4.14                       | 3.09                       | 3.23                               | 14.34                              |
| <b>POLMONE</b>            | 5.70                       | 0.43                       | 0.43                               | 0.39                               | 13.77                      | 6.61                       | 6.71                               | 2.70                               |
| <b>MAMMELLA</b>           | .                          | .                          | .                                  | .                                  | 7.78                       | 2.17                       | 2.13                               | 13.03                              |
| <b>UTERO</b>              | .                          | .                          | .                                  | .                                  | 4.07                       | 4.55                       | 4.46                               | 5.10                               |
| <b>PROSTATA</b>           | 2.42                       | 9.58                       | 9.42                               | 9.71                               | .                          | .                          | .                                  | .                                  |
| <b>LEUCEMIA</b>           | 1.51                       | 8.45                       | 8.49                               | 15.24                              | 2.20                       | 12.40                      | 12.55                              | 27.52                              |
| <b>TUTTI I<br/>TUMORI</b> | 0.61                       | 5.21                       | 5.13                               | 2.84                               | 3.70                       | 2.16                       | 2.04                               | 7.51                               |
| <b>MEDIANA<br/>TOTALE</b> | 3.49                       | 5.21                       | 5.13                               | 6.36                               | 4.11                       | 3.82                       | 3.85                               | 9.55                               |
| <b>AAPD &lt; 5%</b>       | 5                          | 3                          | 3                                  | 3                                  | 5                          | 5                          | 5                                  | 1                                  |

**Tabella 13:** Media della deviazione percentuale assoluta per i tassi di mortalità previsti secondo i modelli lineare, Poisson, log-lineare e AP bayesiano per ciascun tumore per maschi e femmine per il periodo 2005-09. Mediana degli AAPD per ciascun modello di previsione. Numero di AAPD <5% per ciascun modello.

In conclusione, la Tabella 14 mostra le mediane degli AAPD per tutti modelli per il periodo 2000-2009.

|                | <i>Mediana AAPD</i>    |                        |                                 |                                 |
|----------------|------------------------|------------------------|---------------------------------|---------------------------------|
|                | <b>Modello Lineare</b> | <b>Modello Poisson</b> | <b>Modello Log-<br/>Lineare</b> | <b>Modello AP<br/>Bayesiano</b> |
| <b>TOTALE</b>  | 2.68                   | 3.24                   | 3.19                            | 5.36                            |
| <b>Maschi</b>  | 2.62                   | 3.24                   | 3.19                            | 3.98                            |
| <b>Femmine</b> | 2.75                   | 2.84                   | 2.84                            | 7.10                            |

**Tabella 14:** Mediana degli AAPD per ciascun modello di previsione per il periodo 2000-09.

Nel complesso, la mediana degli AAPD del modello lineare risulta minore di quella degli altri modelli, soprattutto di quello AP bayesiano. Si deduce che il modello AP bayesiano sia meno adeguato a fare previsioni di breve-medio periodo rispetto agli altri tre modelli.

Osservando le mediane separatamente per i due sessi, si nota che nei maschi il modello lineare risulta essere il migliore, mentre nelle femmine il modello bayesiano risulta essere il peggiore con mediana maggiore rispetto agli altri tre modelli, i quali presentano mediane simili.

Anche nel caso delle previsioni svolte con periodi annuali, il modello lineare sembra predire meglio i dati di mortalità rispetto agli altri modelli.



# CAPITOLO 6

## Conclusioni

---

L'analisi età-periodo-coorte è uno strumento estremamente utile per lo studio dei dati di mortalità, in particolare per lo studio degli effetti di coorte, ma deve essere usato con la dovuta cautela per evitare di essere indotti verso conclusioni sbagliate.

In mancanza di un'analisi di comparazione simulativa formale tra i modelli in oggetto, dobbiamo limitarci a constatare che i due modelli basati sulla distribuzione della deriva (metodo della verosimiglianza penalizzata e metodo dei GAM) sono maggiormente indicati per confronti epidemiologici, dove gli effetti dell'età sono ben definiti (come nel caso della mortalità per tumori) e le problematiche maggiori risiedono in maniera preponderante nello sbrogliare l'effetto deriva dalle coorti e dai periodi. Mentre ci chiediamo se sia possibile che il modello dei PLS si riveli un metodo di predizione dei dati futuri più performante.

Per quanto concerne l'analisi predittiva, dai risultati ottenuti dal confronto applicativo su un dataset di dimensioni relativamente piccole, dobbiamo limitarci a constatare che il modello lineare e il modello età-periodo-coorte bayesiano sembrerebbero fornire previsioni della mortalità migliori rispetto a quelle degli altri modelli quando si ha una mortalità bassa, mentre nel caso di un maggiore numero di casi sembrano più performanti i modelli di Poisson e log-lineare. I modelli che predicono meglio dipendono dalla forma della distribuzione sconosciuta dei dati analizzati. Tuttavia, tutti i modelli considerati risultano essere più appropriati per predire dati di breve periodo. Nessuno dei modelli considerati ha una buona performance per il medio periodo. Ad oggi la letteratura statistica, in ambito di predizione della mortalità, pone l'attenzione quasi esclusivamente su breve ( $\leq 5$  anni) e lungo ( $\geq 25$  anni) periodo, lasciando scoperto il medio periodo. Attualmente il ricercatore che si occupa di predizioni della mortalità si trova ad affrontare un problema di soggettività nella scelta del modello quando l'orizzonte temporale di previsione è compreso circa tra i 5 e 25 anni. Pertanto, in futuro vorrei continuare ad

occuparmi delle metodologie statistiche riguardanti lo studio degli andamenti della mortalità, in particolare dei metodi di previsione nel medio periodo. Bisogna tenere presente che la previsione dei trend futuri è sempre un esercizio complesso e insidioso, e le stime ottenute dovrebbero essere prese con la dovuta cautela e considerate solo come indicazioni generali di potenziale interesse per l'epidemiologia e la programmazione sanitaria.



# Bibliografia

---

1. Frova, L., S. Marchetti, and M. Pace, *La codifica automatica delle cause di morte in Italia*. 2005: Istituto Nazionale di Statistica
2. Esteve, J., E. Benhamou, and L. Raymond, *Statistical methods in cancer research. Volume IV. Descriptive epidemiology*. IARC Sci Publ, 1994(128): p. 1-302.
3. Holford, T., *Age-Period-Cohort Analysis*, in *Encyclopedia of Statistics in Behavioral Science*. 2005, John Wiley & Sons, Ltd.
4. Malvezzi, M., et al., *European cancer mortality predictions for the year 2011*. Ann Oncol, 2011. **22**(4): p. 947-56.
5. Zocchetti, C. and D. Consonni, *[Mortality rate and its statistical properties]*. Med Lav, 1994. **85**(4): p. 327-43.
6. Isabel dos Santos Silva, *Cancer epidemiology: principles and methods*, ed. IARC. 1999, Lyon, France.
7. Doll, R. and P.G. Smith, *Comparison between registries: age-standardized rates. Vol. IV. IARC Sci Publ No. 42*, in *Cancer Incidence in Five Continents*, Waterhouse JAH, et al., Editors. 1982, IARC: Lyon. p. 671-675.
8. Brillinger, D.R., *The natural variability of vital rates and associated statistics*. Biometrics, 1986. **42**(4): p. 693-734.
9. Kim, H.J., et al., *Permutation tests for joinpoint regression with applications to cancer rates*. Stat Med, 2000. **19**(3): p. 335-51.
10. Frost, W.H., *The age selection of mortality from tuberculosis in successive decades*. American Journal of Hygiene, 1939. **Section A**(30): p. 91-96.
11. Osmond, C. and M.J. Gardner, *Age, period and cohort models applied to cancer mortality rates*. Stat Med, 1982. **1**(3): p. 245-59.
12. Decarli, A. and C. La Vecchia, *Age, period and cohort models: review of knowledge and implementation in GLIM*. Rivista Statistica Applicata, 1987. **20**: p. 397-410.
13. Holford, T.R., *The estimation of age, period and cohort effects for vital rates*. Biometrics, 1983. **39**(2): p. 311-24.
14. Malvezzi, M., P. Bertuccio, and V. Edefonti, *Age period cohort analysis of cancer mortality data: methods and application to italian male mortality data for gastric cancer and cancers of the oral cavity and pharynx*. BioMedical Statistics and Clinical Epidemiology, 2009. **3**(2): p. 113-122.
15. Frost, W.H., *The age selection of mortality from tuberculosis in successive decades. 1939*. Am J Epidemiol, 1995. **141**(1): p. 4-9; discussion 3.
16. Carstensen, B., *Age-period-cohort models for the Lexis diagram*. Stat Med, 2007. **26**(15): p. 3018-45.
17. Kupper, L.L., et al., *Statistical age-period-cohort analysis: a review and critique*. J Chronic Dis, 1985. **38**(10): p. 811-30.
18. Clayton, D. and E. Schifflers, *Models for temporal variation in cancer rates. II: Age-period-cohort models*. Stat Med, 1987. **6**(4): p. 469-81.

19. Holford, T.R., Z. Zhang, and L.A. McKay, *Estimating age, period and cohort effects using the multistage model for cancer*. Stat Med, 1994. **13**(1): p. 23-41.
20. Robertson, C. and P. Boyle, *Age-period-cohort analysis of chronic disease rates. I: Modelling approach*. Stat Med, 1998. **17**(12): p. 1305-23.
21. Held, L., Schmid, V., *Bayesian age-period-cohort modeling and prediction - BAMP*. J Stat Softw, 2007. **21**(8).
22. Smith, H.L., *Advances in age-period-cohort analysis*. Sociological Methods Research, 2008. **36**: p. 287-296.
23. Yang, Y., Schulhofer-Wohl, S., Fu, W., Land, K., *The intrinsic estimator for age-period-cohort analysis. What it is and how to use it*. AJS, 2008. **113**: p. 1697-1736.
24. Fukuda, K., *Age-period-cohort decomposition using principal components or partial least squares: a simulation study*, in 2009, College of Economics, Nihon University: Tokyo.
25. Clayton, D. and E. Schifflers, *Models for temporal variation in cancer rates. I: Age-period and age-cohort models*. Stat Med, 1987. **6**(4): p. 449-67.
26. Efron, B., Tibshirani R, *An introduction to the bootstrap*. 1993.
27. Carstensen, B., Keiding, N., *Age-Period-Cohort models: Statistical inference in the Lexis diagram*, in *Department of Biostatistics Institute of Public Health*. 2005, University of Copenhagen.
28. Tu, Y.K., G. Davey Smith, and M.S. Gilthorpe, *A new approach to age-period-cohort analysis using partial least squares regression: the trend in blood pressure in the Glasgow Alumni cohort*. PLoS One, 2011. **6**(4): p. e19401.
29. Tu, Y.K., N. Kramer, and W.C. Lee, *Addressing the identification problem in age-period-cohort analysis: a tutorial on the use of partial least squares and principal components analysis*. Epidemiology, 2012. **23**(4): p. 583-93.
30. Zhou, X., Obuchowski, N., McClish, D., *Appendix B: Jackknife and Bootstrap Methods of Estimating Variances and Confidence Intervals*, in *Statistical Methods in Diagnostic Medicine*. 2011.
31. Qiu Z., J.Z., Wang M., Hatcher J. et al, *Long-term projection methods: Comparison of age-period-cohort model based approaches*. 2010, Alberta Health Services for the Canadian Partnership against Cancer.
32. Moller, B., et al., *Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches*. Stat Med, 2003. **22**(17): p. 2751-66.
33. Lee, T.C., C.B. Dean, and R. Semenciw, *Short-term cancer mortality projections: a comparative study of prediction methods*. Stat Med, 2011. **30**(29): p. 3387-402.
34. Clements, M.S., B.K. Armstrong, and S.H. Moolgavkar, *Lung cancer rate predictions using generalized additive models*. Biostatistics, 2005. **6**(4): p. 576-89.
35. Lerman, P.M., *Fitting segmented regression models by grid search*. Appl Statist, 1980. **29**(1): p. 77-84.
36. Faraway, J.J., *Linear models with R*. 2009: Chapman & Hall, Boca Raton.
37. Bray, I., *Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality*. Appl. Statist., 2002. **51**: p. 151-164.
38. Berzuini, C. and D. Clayton, *Bayesian analysis of survival on multiple time scales*. Stat Med, 1994. **13**(8): p. 823-38.

- 
39. Holford, T.R., *An alternative approach to statistical age-period-cohort analysis*. J Chronic Dis, 1985. **38**(10): p. 831-40.
  40. Breslow, N.E. and D.G. Clayton, *Approximate inference in generalized linear mixed models*. J. Am. Statist. Ass., 1993. **88**: p. 9-25.
  41. Plummer, M., et al., *CODA: Convergence Diagnosis and Output Analysis for MCMC*. R News, 2006. **6**: p. 7-11.
  42. Qiu, Z., Z. Jiang, and J. Hatcher, *Comparison of short-term projection methods: Proposing the hybrid approach and bayesian models*. 2010, Alberta Health Services for the Canadian Partnership against Cancer. p. 1-70.
  43. World Health Organization Statistical Information System, *WHO mortality database*. Available at: <http://www3.who.int/whosis/menu.cfm>, Last accessed November 2011.
  44. World Health Organization, *International Classification of Disease: 8th revision*. 1967, Geneva: World Health Organization.
  45. World Health Organization, *International Classification of Disease: 9th revision*. 1977, Geneva.
  46. World Health Organization, *International statistical classification of disease and related health problems, 10th revision*, ed. W.H. Organization. 1992, Geneva.
  47. Anderson, R.N., et al., *Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates*. Natl Vital Stat Rep, 2001. **49**(2): p. 1-32.
  48. Anderson, R.N. and H.M. Rosenberg, *Disease classification: measuring the effect of the Tenth Revision of the International Classification of Diseases on cause-of-death data in the United States*. Stat Med, 2003. **22**(9): p. 1551-70.
  49. Malvezzi, M., et al., *Lung cancer mortality in European men: trends and predictions*. Lung Cancer, 2013. **80**(2): p. 138-45.
  50. World Health Organization, *Tobacco or Health: a global status report*. 1997, Geneva, Switzerland: WHO.
  51. World Health Organization Statistical Information System. *Global Adult Tobacco Survey*. Available at: <http://www.who.int/tobacco/surveillance/gats/en/index.html> (accessed 2011). 2011.
  52. Karim-Kos, H.E., et al., *Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s*. Eur J Cancer, 2008. **44**(10): p. 1345-89.



# Appendice 1

---

Per il calcolo dei tassi dell'UE sono stati usati i dati dei 27 Stati membri sanciti dal trattato ufficiale del 1° gennaio 2007 (Austria, Belgio, Bulgaria, Cipro, Danimarca, Estonia, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lettonia, Lituania, Lussemburgo, Malta, Paesi Bassi, Polonia, Portogallo, Regno Unito, Repubblica Ceca, Romania, Slovacchia, Slovenia, Spagna, Svezia, Ungheria), con l'eccezione di Cipro per dati insufficienti. Quando i dati non erano disponibili per un Paese, sono stati replicati i dati disponibili più vicini (ovvero, generalmente quello dell'anno precedente o successivo) come segue.

Per i tumori dello stomaco, della mammella, della prostata e per tutti i tumori: per il Belgio i dati degli anni 2003-04 sono stati usati per il periodo 2000-02; per la Germania quelli del 1973 sono stati usati per tutto il quinquennio 1970-74 e quelli del 1978 per il 1979; per l'Italia quelli del 2003 sono stati usati per il 2004 e quelli del 2006 per il 2005; per la Lettonia i dati del quinquennio 1980-84 sono stati replicati per il 1970-74 e 1975-79; per il Lussemburgo quelli del 1971 per il 1970; per la Polonia quelli del 1996 per il 1997 e quelli del 1999 per il 1998; per il Portogallo quelli del 1971 per il 1970; per la Romania i dati del 1978 sono stati usati per il 1979; per la Slovenia quelli del quinquennio 1985-89 sono stati replicati per i quinquenni 1970-74, 1975-79 e 1980-84; per la Spagna quelli del 1971 per il 1970; per l'Estonia una media dei dati del 1981 e 1982 è stata usata per il quinquennio 1980-84, la quale è stata poi replicata per i quinquenni 1970-74 e 1975-79, lo stesso procedimento è stato applicato al dataset della Lituania; i dati per la Cecoslovacchia sono stati usati per il periodo 1970-1991, mentre dal 1992 in avanti sono stati usati i dati separatamente per Repubblica Ceca e Slovacchia.

Per i tumori dell'intestino (colon e retto), del polmone, dell'utero (cervice e corpo) e la leucemia, oltre alle operazioni descritte sopra: per il Portogallo i dati di mortalità erano mancanti tra il 2004 e il 2006, pertanto i dati del 2003 sono stati replicati per tutto il periodo 2003-07.



Per il tumore del pancreas, oltre alle operazioni svolte per i dati di mortalità per tutti i tumori, sono state fatte anche le seguenti replicazioni: per il Belgio i dati del 1976 sono stati usati per il 1977 e quelli del 1979 per il 1978; per l'Estonia si è calcolata la media dei dati dal 1993 al 1996 e replicata per tutto il periodo 1970-1992; per la Lettonia la media dei dati del periodo 1996-1999 è stata replicata per il 1970-1996; per la Lituania la media dei dati del periodo 1993-96 è stata replicata per gli anni dal 1970 al 1995; per il Lussemburgo, il Portogallo e la Romania la media dei dati del quinquennio 1980-84 è stata applicata al decennio 1970-79; per Malta la media dei dati del 1978-80 è stata applicata agli anni tra il 1970 e il 1977; per la Polonia la media del quinquennio 1980-84 è stata replicata per quello del 1975-79, mentre quella del periodo 1970-72 è stata replicata per il 1973-74.

## Appendice 2

---

Di seguito vengono riportate le tabelle per i tassi di mortalità osservati e previsti con il metodo età-periodo con joinpoint sul numero di morti e con i metodi età-periodo e età-periodo-coorte bayesiani per i principali siti tumorali per il periodo 2000-2009 (periodi quinquennali e annuali) nell'UE.

– Periodi quinquennali:

| TUMORE DELLO STOMACO |      |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
|----------------------|------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|---------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso                |      |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
| Maschi               |      |                             |                             |                                     |                                     |                                      | Femmine |                             |                             |                                     |                                     |                                      |
|                      | Oss. | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |         | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| <b>2000-04</b>       | 9.62 | 9.31                        | 9.65                        | 9.64                                | 9.65                                | 9.69                                 | 4.39    | 4.17                        | 4.36                        | 4.35                                | 4.31                                | 4.47                                 |
| <b>2005-09</b>       | 8.00 | 7.29                        | 8.17                        | 8.16                                | 8.04                                | 8.11                                 | 3.65    | 3.22                        | 3.74                        | 3.73                                | 3.57                                | 3.82                                 |

| TUMORE DELL'INTESTINO (COLON E RETTO) |       |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
|---------------------------------------|-------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|---------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso                                 |       |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
| Maschi                                |       |                             |                             |                                     |                                     |                                      | Femmine |                             |                             |                                     |                                     |                                      |
|                                       | Oss.  | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |         | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| <b>2000-04</b>                        | 18.90 | 19.20                       | 19.53                       | 19.51                               | 19.57                               | 19.37                                | 11.46   | 11.51                       | 11.74                       | 11.74                               | 11.98                               | 11.71                                |
| <b>2005-09</b>                        | 17.73 | 18.34                       | 19.49                       | 19.45                               | 19.43                               | 18.87                                | 10.51   | 10.52                       | 11.36                       | 11.35                               | 11.35                               | 10.95                                |

| TUMORE DEL PANCREAS |      |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
|---------------------|------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|---------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso               |      |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
| Maschi              |      |                             |                             |                                     |                                     |                                      | Femmine |                             |                             |                                     |                                     |                                      |
|                     | Oss. | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |         | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| <b>2000-04</b>      | 7.61 | 7.55                        | 7.67                        | 7.67                                | 7.50                                | 7.44                                 | 4.97    | 4.94                        | 5.04                        | 5.04                                | 4.96                                | 4.85                                 |
| <b>2005-09</b>      | 7.84 | 7.46                        | 7.86                        | 7.87                                | 7.46                                | 7.32                                 | 5.24    | 5.02                        | 5.40                        | 5.41                                | 5.06                                | 4.89                                 |

| TUMORE DEL POLMONE |       |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
|--------------------|-------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|---------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso              |       |                             |                             |                                     |                                     |                                      |         |                             |                             |                                     |                                     |                                      |
| Maschi             |       |                             |                             |                                     |                                     |                                      | Femmine |                             |                             |                                     |                                     |                                      |
|                    | Oss.  | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |         | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| <b>2000-04</b>     | 44.69 | 43.76                       | 44.42                       | 44.42                               | 46.57                               | 46.07                                | 11.30   | 10.69                       | 10.94                       | 10.93                               | 11.42                               | 11.08                                |
| <b>2005-09</b>     | 40.92 | 38.57                       | 40.82                       | 40.83                               | 43.99                               | 42.92                                | 12.63   | 10.88                       | 11.79                       | 11.78                               | 12.39                               | 11.92                                |

| <b>TUMORE DELLA MAMMELLA</b> |             |                                      |                                      |  |   |  |
|------------------------------|-------------|--------------------------------------|--------------------------------------|--|---|--|
| <b>Sesso</b>                 |             |                                      |                                      |  |   |  |
| <b>Femmine</b>               |             |                                      |                                      |  |   |  |
|                              | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> | <b>Prev.<br/>Modello APC<br/>Bayesiano</b> |
| <b>2000-04</b>               | 17.94       | 17.34                                | 17.60                                | 17.60                                    | 19.23                                     | 18.39                                      |
| <b>2005-09</b>               | 16.43       | 15.15                                | 16.07                                | 16.07                                    | 18.45                                     | 17.11                                      |

| <b>TUMORE DELL'UTERO (CERVICE E CORPO)</b> |             |                                      |                                      |  |   |  |
|--|-------------|--------------------------------------|--------------------------------------|--|---|--|
| <b>Sesso</b>                               |             |                                      |                                      |  |   |  |
| <b>Femmine</b>                             |             |                                      |                                      |  |   |  |
|  | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> | <b>Prev.<br/>Modello APC<br/>Bayesiano</b> |
| <b>2000-04</b>                             | 5.61        | 5.59                                 | 5.73                                 | 5.72                                     | 5.82                                      | 5.63                                       |
| <b>2005-09</b>                             | 5.15        | 4.94                                 | 5.38                                 | 5.38                                     | 5.25                                      | 5.10                                       |

| TUMORE DELLA PROSTATA |       |                             |                             |                                 |                                  |                                   |
|-----------------------|-------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|-----------------------------------|
| Sesso                 |       |                             |                             |                                 |                                  |                                   |
| Femmine               |       |                             |                             |                                 |                                  |                                   |
|                       | Oss.  | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano | Prev.<br>Modello APC<br>Bayesiano |
| <b>2000-04</b>        | 13.75 | 13.79                       | 14.05                       | 14.04                           | 14.69                            | 14.67                             |
| <b>2005-09</b>        | 12.18 | 12.47                       | 13.34                       | 13.32                           | 14.65                            | 14.63                             |

| LEUCEMIA       |      |                             |                             |                                     |                                     |                                      |      |                             |                             |                                     |                                     |                                      |
|----------------|------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|------|-----------------------------|-----------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Sesso          |      |                             |                             |                                     |                                     |                                      |      |                             |                             |                                     |                                     |                                      |
| Maschi         |      |                             |                             |                                     |                                     | Femmine                              |      |                             |                             |                                     |                                     |                                      |
|                | Oss. | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano | Oss. | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-<br>lineare | Prev.<br>Modello<br>AP<br>Bayesiano | Prev.<br>Modello<br>APC<br>Bayesiano |
| <b>2000-04</b> | 5.22 | 5.11                        | 5.32                        | 5.32                                | 5.58                                | 5.10                                 | 3.20 | 3.08                        | 3.28                        | 3.28                                | 3.74                                | 3.19                                 |
| <b>2005-09</b> | 4.85 | 4.77                        | 5.26                        | 5.26                                | 5.44                                | 4.77                                 | 2.91 | 2.85                        | 3.27                        | 3.28                                | 3.70                                | 2.99                                 |

– Periodi annuali:

| <b>TUMORE DELLO STOMACO</b> |             |                                      |                                      |  |   |                |                                      |                                      |  |   |
|-----------------------------|-------------|--------------------------------------|--------------------------------------|--|---|----------------|--------------------------------------|--------------------------------------|--|---|
| <b>Sesso</b>                |             |                                      |                                      |  |   |                |                                      |                                      |  |   |
| <b>Maschi</b>               |             |                                      |                                      |  |   | <b>Femmine</b> |                                      |                                      |  |   |
|                             | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |                | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |
| <b>2000</b>                 | 10.33       | 10.17                                | 10.32                                | 10.31                                    | 10.44                                     | 4.71           | 4.57                                 | 4.66                                 | 4.65                                     | 4.57                                      |
| <b>2001</b>                 | 9.94        | 9.72                                 | 9.96                                 | 9.95                                     | 10.11                                     | 4.51           | 4.37                                 | 4.49                                 | 4.49                                     | 4.37                                      |
| <b>2002</b>                 | 9.68        | 9.32                                 | 9.64                                 | 9.63                                     | 9.80                                      | 4.37           | 4.17                                 | 4.35                                 | 4.34                                     | 4.18                                      |
| <b>2003</b>                 | 9.23        | 8.94                                 | 9.37                                 | 9.36                                     | 9.51                                      | 4.23           | 4.00                                 | 4.23                                 | 4.22                                     | 4.01                                      |
| <b>2004</b>                 | 8.97        | 8.55                                 | 9.09                                 | 9.08                                     | 9.23                                      | 4.12           | 3.82                                 | 4.12                                 | 4.11                                     | 3.84                                      |
| <b>2005</b>                 | 8.57        | 8.07                                 | 8.73                                 | 8.72                                     | 8.97                                      | 3.88           | 3.61                                 | 3.97                                 | 3.96                                     | 3.69                                      |
| <b>2006</b>                 | 8.27        | 7.65                                 | 8.44                                 | 8.43                                     | 8.73                                      | 3.74           | 3.41                                 | 3.85                                 | 3.84                                     | 3.55                                      |
| <b>2007</b>                 | 7.98        | 7.27                                 | 8.15                                 | 8.14                                     | 8.52                                      | 3.62           | 3.20                                 | 3.73                                 | 3.72                                     | 3.41                                      |
| <b>2008</b>                 | 7.76        | 6.95                                 | 7.91                                 | 7.90                                     | 8.32                                      | 3.56           | 3.04                                 | 3.63                                 | 3.62                                     | 3.29                                      |
| <b>2009</b>                 | 7.48        | 6.62                                 | 7.67                                 | 7.66                                     | 8.14                                      | 3.44           | 2.89                                 | 3.53                                 | 3.52                                     | 3.18                                      |

| TUMORE DELL'INTESTINO (COLON E RETTO) |        |                             |                             |                                 |                                  |       |                             |                             |                                 |                                  |
|---------------------------------------|--------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|-------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|
| Sesso                                 |        |                             |                             |                                 |                                  |       |                             |                             |                                 |                                  |
|                                       | Maschi |                             |                             |                                 |                                  |       | Femmine                     |                             |                                 |                                  |
|                                       | Oss.   | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano |       | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano |
| <b>2000</b>                           | 19.04  | 19.40                       | 19.52                       | 19.50                           | 19.46                            | 11.79 | 11.87                       | 11.94                       | 11.94                           | 12.26                            |
| <b>2001</b>                           | 19.09  | 19.28                       | 19.48                       | 19.46                           | 19.32                            | 11.62 | 11.67                       | 11.81                       | 11.81                           | 12.16                            |
| <b>2002</b>                           | 19.03  | 19.20                       | 19.51                       | 19.49                           | 19.19                            | 11.46 | 11.51                       | 11.72                       | 11.72                           | 12.05                            |
| <b>2003</b>                           | 18.78  | 19.17                       | 19.60                       | 19.58                           | 19.07                            | 11.39 | 11.39                       | 11.69                       | 11.69                           | 11.96                            |
| <b>2004</b>                           | 18.56  | 19.09                       | 19.67                       | 19.65                           | 18.97                            | 11.07 | 11.25                       | 11.66                       | 11.65                           | 11.88                            |
| <b>2005</b>                           | 18.16  | 18.77                       | 19.50                       | 19.47                           | 18.90                            | 10.87 | 10.99                       | 11.52                       | 11.51                           | 11.81                            |
| <b>2006</b>                           | 17.95  | 18.58                       | 19.50                       | 19.47                           | 18.85                            | 10.58 | 10.76                       | 11.43                       | 11.43                           | 11.75                            |
| <b>2007</b>                           | 17.59  | 18.34                       | 19.45                       | 19.42                           | 18.83                            | 10.51 | 10.52                       | 11.33                       | 11.33                           | 11.71                            |
| <b>2008</b>                           | 17.57  | 18.16                       | 19.50                       | 19.46                           | 18.84                            | 10.43 | 10.30                       | 11.29                       | 11.29                           | 11.68                            |
| <b>2009</b>                           | 17.43  | 17.95                       | 19.55                       | 19.50                           | 18.90                            | 10.16 | 10.07                       | 11.26                       | 11.26                           | 11.68                            |



| TUMORE DEL PANCREAS |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|---------------------|--------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|---------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|
| Sesso               |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|                     | Maschi |                             |                             |                                 |                                  | Femmine |                             |                             |                                 |                                  |
|                     | Oss.   | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano | Oss.    | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano |
| <b>2000</b>         | 7.56   | 7.55                        | 7.59                        | 7.59                            | 7.63                             | 4.93    | 4.89                        | 4.92                        | 4.92                            | 5.06                             |
| <b>2001</b>         | 7.59   | 7.55                        | 7.62                        | 7.62                            | 7.66                             | 4.98    | 4.91                        | 4.97                        | 4.97                            | 5.15                             |
| <b>2002</b>         | 7.60   | 7.55                        | 7.66                        | 7.66                            | 7.71                             | 4.96    | 4.94                        | 5.03                        | 5.03                            | 5.24                             |
| <b>2003</b>         | 7.62   | 7.57                        | 7.72                        | 7.72                            | 7.77                             | 5.01    | 4.98                        | 5.11                        | 5.11                            | 5.35                             |
| <b>2004</b>         | 7.66   | 7.57                        | 7.78                        | 7.78                            | 7.84                             | 5.00    | 5.02                        | 5.20                        | 5.20                            | 5.47                             |
| <b>2005</b>         | 7.76   | 7.52                        | 7.77                        | 7.78                            | 7.93                             | 5.09    | 5.02                        | 5.25                        | 5.25                            | 5.61                             |
| <b>2006</b>         | 7.78   | 7.50                        | 7.82                        | 7.83                            | 8.04                             | 5.16    | 5.02                        | 5.32                        | 5.32                            | 5.77                             |
| <b>2007</b>         | 7.85   | 7.46                        | 7.85                        | 7.86                            | 8.18                             | 5.31    | 5.02                        | 5.39                        | 5.39                            | 5.96                             |
| <b>2008</b>         | 7.95   | 7.43                        | 7.91                        | 7.92                            | 8.36                             | 5.29    | 5.02                        | 5.48                        | 5.49                            | 6.18                             |
| <b>2009</b>         | 7.86   | 7.40                        | 7.97                        | 7.98                            | 8.57                             | 5.34    | 5.03                        | 5.58                        | 5.59                            | 6.44                             |

| TUMORE DEL POLMONE |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|--------------------|--------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|---------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|
| Sesso              |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|                    | Maschi |                             |                             |                                 |                                  | Femmine |                             |                             |                                 |                                  |
|                    | Oss.   | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano | Oss.    | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano |
| <b>2000</b>        | 46.25  | 45.87                       | 46.09                       | 46.09                           | 46.71                            | 10.96   | 10.58                       | 10.67                       | 10.66                           | 10.99                            |
| <b>2001</b>        | 45.47  | 44.79                       | 45.18                       | 45.19                           | 45.71                            | 10.99   | 10.64                       | 10.79                       | 10.78                           | 11.13                            |
| <b>2002</b>        | 44.69  | 43.77                       | 44.38                       | 44.38                           | 44.76                            | 11.34   | 10.69                       | 10.92                       | 10.91                           | 11.29                            |
| <b>2003</b>        | 43.78  | 42.82                       | 43.68                       | 43.68                           | 43.85                            | 11.51   | 10.75                       | 11.08                       | 11.07                           | 11.45                            |
| <b>2004</b>        | 43.35  | 41.90                       | 43.04                       | 43.05                           | 43.00                            | 11.69   | 10.82                       | 11.27                       | 11.26                           | 11.63                            |
| <b>2005</b>        | 42.27  | 40.65                       | 42.11                       | 42.11                           | 42.21                            | 12.10   | 10.83                       | 11.40                       | 11.39                           | 11.82                            |
| <b>2006</b>        | 41.71  | 39.68                       | 41.49                       | 41.50                           | 41.49                            | 12.35   | 10.87                       | 11.58                       | 11.57                           | 12.03                            |
| <b>2007</b>        | 40.93  | 38.60                       | 40.79                       | 40.80                           | 40.83                            | 12.64   | 10.88                       | 11.77                       | 11.76                           | 12.26                            |
| <b>2008</b>        | 40.36  | 37.60                       | 40.23                       | 40.23                           | 40.25                            | 12.98   | 10.90                       | 11.98                       | 11.97                           | 12.51                            |
| <b>2009</b>        | 39.44  | 36.58                       | 39.67                       | 39.67                           | 39.74                            | 13.06   | 10.91                       | 12.21                       | 12.19                           | 12.80                            |

| <b>TUMORE DELLA MAMMELLA</b> |             |                                      |                                      |  |   |
|------------------------------|-------------|--------------------------------------|--------------------------------------|--|---|
| <b>Sesso</b>                 |             |                                      |                                      |  |   |
| <b>Femmine</b>               |             |                                      |                                      |  |   |
|                              | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |
| <b>2000</b>                  | 18.55       | 18.22                                | 18.30                                | 18.30                                    | 19.30                                     |
| <b>2001</b>                  | 18.24       | 17.77                                | 17.92                                | 17.93                                    | 19.15                                     |
| <b>2002</b>                  | 17.94       | 17.34                                | 17.58                                | 17.58                                    | 19.01                                     |
| <b>2003</b>                  | 17.58       | 16.95                                | 17.29                                | 17.29                                    | 18.88                                     |
| <b>2004</b>                  | 17.43       | 16.56                                | 17.02                                | 17.02                                    | 18.76                                     |
| <b>2005</b>                  | 17.04       | 16.08                                | 16.66                                | 16.67                                    | 18.67                                     |
| <b>2006</b>                  | 16.72       | 15.63                                | 16.36                                | 16.37                                    | 18.59                                     |
| <b>2007</b>                  | 16.29       | 15.15                                | 16.05                                | 16.05                                    | 18.54                                     |
| <b>2008</b>                  | 16.28       | 14.70                                | 15.78                                | 15.79                                    | 18.51                                     |
| <b>2009</b>                  | 15.84       | 14.25                                | 15.53                                | 15.54                                    | 18.52                                     |

| <b>TUMORE DELL'UTERO (CERVICE E CORPO)</b> |             |                                      |                                      |  |   |
|--|-------------|--------------------------------------|--------------------------------------|--|---|
| <b>Sesso</b>                               |             |                                      |                                      |  |   |
| <b>Femmine</b>                             |             |                                      |                                      |  |   |
|  | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |
| <b>2000</b>                                | 5.91        | 5.85                                 | 5.91                                 | 5.90                                     | 6.08                                      |
| <b>2001</b>                                | 5.70        | 5.71                                 | 5.81                                 | 5.80                                     | 5.96                                      |
| <b>2002</b>                                | 5.59        | 5.59                                 | 5.72                                 | 5.72                                     | 5.85                                      |
| <b>2003</b>                                | 5.48        | 5.47                                 | 5.65                                 | 5.65                                     | 5.74                                      |
| <b>2004</b>                                | 5.40        | 5.35                                 | 5.59                                 | 5.59                                     | 5.64                                      |
| <b>2005</b>                                | 5.28        | 5.21                                 | 5.50                                 | 5.50                                     | 5.55                                      |
| <b>2006</b>                                | 5.19        | 5.08                                 | 5.44                                 | 5.43                                     | 5.47                                      |
| <b>2007</b>                                | 5.18        | 4.94                                 | 5.37                                 | 5.37                                     | 5.40                                      |
| <b>2008</b>                                | 5.08        | 4.81                                 | 5.32                                 | 5.32                                     | 5.34                                      |
| <b>2009</b>                                | 5.02        | 4.68                                 | 5.28                                 | 5.28                                     | 5.30                                      |

| <b>TUMORE DELLA PROSTATA</b> |             |                                      |                                      |  |   |
|------------------------------|-------------|--------------------------------------|--------------------------------------|--|---|
| <b>Sesso</b>                 |             |                                      |                                      |  |   |
| <b>Femmine</b>               |             |                                      |                                      |  |   |
|                              | <b>Oss.</b> | <b>Prev.<br/>Modello<br/>Lineare</b> | <b>Prev.<br/>Modello<br/>Poisson</b> | <b>Prev.<br/>Modello<br/>Log-lineare</b> | <b>Prev.<br/>Modello AP<br/>Bayesiano</b> |
| <b>2000</b>                  | 13.96       | 14.12                                | 14.21                                | 14.20                                    | 14.26                                     |
| <b>2001</b>                  | 13.89       | 13.91                                | 14.06                                | 14.05                                    | 14.08                                     |
| <b>2002</b>                  | 13.79       | 13.80                                | 14.04                                | 14.03                                    | 13.92                                     |
| <b>2003</b>                  | 13.74       | 13.76                                | 14.10                                | 14.09                                    | 13.77                                     |
| <b>2004</b>                  | 13.41       | 13.62                                | 14.07                                | 14.06                                    | 13.63                                     |
| <b>2005</b>                  | 12.73       | 13.17                                | 13.74                                | 13.73                                    | 13.51                                     |
| <b>2006</b>                  | 12.47       | 12.83                                | 13.54                                | 13.52                                    | 13.42                                     |
| <b>2007</b>                  | 12.17       | 12.46                                | 13.31                                | 13.29                                    | 13.35                                     |
| <b>2008</b>                  | 11.95       | 12.17                                | 13.19                                | 13.17                                    | 13.31                                     |
| <b>2009</b>                  | 11.69       | 11.87                                | 13.05                                | 13.03                                    | 13.29                                     |

| LEUCEMIA    |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|-------------|--------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|---------|-----------------------------|-----------------------------|---------------------------------|----------------------------------|
| Sesso       |        |                             |                             |                                 |                                  |         |                             |                             |                                 |                                  |
|             | Maschi |                             |                             |                                 |                                  | Femmine |                             |                             |                                 |                                  |
|             | Oss.   | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano | Oss.    | Prev.<br>Modello<br>Lineare | Prev.<br>Modello<br>Poisson | Prev.<br>Modello<br>Log-lineare | Prev.<br>Modello AP<br>Bayesiano |
| <b>2000</b> | 5.28   | 5.22                        | 5.33                        | 5.33                            | 5.66                             | 3.24    | 3.18                        | 3.30                        | 3.30                            | 3.75                             |
| <b>2001</b> | 5.34   | 5.17                        | 5.32                        | 5.32                            | 5.64                             | 3.23    | 3.13                        | 3.28                        | 3.29                            | 3.74                             |
| <b>2002</b> | 5.19   | 5.12                        | 5.32                        | 5.32                            | 5.62                             | 3.22    | 3.08                        | 3.28                        | 3.28                            | 3.72                             |
| <b>2003</b> | 5.18   | 5.07                        | 5.33                        | 5.33                            | 5.60                             | 3.18    | 3.04                        | 3.28                        | 3.28                            | 3.71                             |
| <b>2004</b> | 5.12   | 5.02                        | 5.33                        | 5.34                            | 5.58                             | 3.12    | 3.00                        | 3.29                        | 3.29                            | 3.70                             |
| <b>2005</b> | 4.95   | 4.91                        | 5.28                        | 5.28                            | 5.57                             | 3.00    | 2.93                        | 3.27                        | 3.28                            | 3.70                             |
| <b>2006</b> | 4.96   | 4.85                        | 5.27                        | 5.27                            | 5.57                             | 2.97    | 2.89                        | 3.27                        | 3.27                            | 3.70                             |
| <b>2007</b> | 4.80   | 4.77                        | 5.25                        | 5.25                            | 5.58                             | 2.92    | 2.85                        | 3.27                        | 3.27                            | 3.71                             |
| <b>2008</b> | 4.87   | 4.71                        | 5.24                        | 5.25                            | 5.59                             | 2.86    | 2.81                        | 3.27                        | 3.28                            | 3.72                             |
| <b>2009</b> | 4.66   | 4.65                        | 5.24                        | 5.25                            | 5.61                             | 2.82    | 2.77                        | 3.29                        | 3.29                            | 3.74                             |